UNIVERSITA' DEGLI STUDI DI MILANO

Facoltà di Medicina e Chirurgia

Dipartimento di Medicina del Lavoro "Clinica del Lavoro L. Devoto"

Sezione di Statistica Medica e Biometria "G.A. Maccacaro"

CORSO DI DOTTORATO DI RICERCA IN STATISTICA BIOMEDICA (D4R)

SETTORE SCIENTIFICO DISCIPLINARE MED/01 – XXIV CICLO

TESI DI DOTTORATO DI RICERCA

FLAVONOIDS, PROANTHOCYANIDINS AND CANCER RISK

Statistical methods to disentangle the effects of flavonoids, their sources and other dietary compounds on cancer risk

Dottoranda: **Marta Rossi**

(Matricola R08210)

Relatore: **Prof. Carlo La Vecchia**

Relatore esterno: **Prof. Dimitrios Trichopoulos**

A.A. 2010/2011

CONTENTS

1. RATIONALE

A diet rich in vegetables and fruit has been associated with a reduced risk of various common cancers, particularly of the respiratory and digestive tracts [1]. With specific focus to a network of Italian case-control studies conducted since the early 1990's (Figures 1 and 2), vegetable intake was inversely related to the risk of several common epithelial cancers: the odds ratios (OR) for digestive tract neoplasms ranged between 0.3 and 0.8 for the highest compared with the lowest levels of vegetable intake. Less consistent inverse relations were observed for some hormone-related neoplasms, such as breast and ovary. High fruit intake was associated to reduced ORs of cancers of the upper digestive tract, stomach, colorectum, and Non-Hodgkin's lymphomas. No material effect, however, was observed for fruit intake on neoplasms of the breast, the female genital tract or the prostate [2]. For digestive tract cancers, population-attributable risks for low intake of vegetables and fruit ranged between 15 and 40% [3].

Further investigations have tried to understand whether such a favorable effect may be attributed to specific vitamins, micronutrients or bioactive compounds contained in plant foods. Flavonoids - a large group of phytochemicals with a similar structure, naturally occurring in vegetables, fruit, and beverages of plant origin - have shown antioxidant, antimutagenic, and antiproliferative properties in vitro [4, 5], and have thus been suggested to have a potential protective effect on common cancers [6-9].

The availability of detailed and reliable food composition tables for flavonoids published by the US Department of Agriculture on their 6 major classes (flavanols, flavanones, flavonols, flavones, anthocyanidins and isoflavones) [10, 11], and, more recently, on a class of polymers of flavanols called proanthocyanidins [12] has allowed epidemiological studies to further investigate the role of flavonoids in cancer etiology from the early 2000's. Intake of various flavonoids has been inversely related to the risk of cancers of the upper aerodigestive tract [13], digestive tract [14, 15], breast

[16, 17], and urogenital tract [18, 19]. No epidemiological study has investigated systematically the relation of dietary proanthocyanidins by their degree of polymerization with cancer risk.

Given the high correlation between dietary factors, it is difficult to disentangle the effect of various factors and detect which specific component or group of components is the responsible for a protective role against cancer. Moreover, since several antioxidants influence cancer risk and may act synergistically against oxidative stress to prevent carcinogenesis, examining overall antioxidant exposure rather than individual antioxidants has recently been suggested [20]. An inverse association between total antioxidant capacity (TAC) and rectal cancer risk has been reported from a large US cohort study [20].

2. OBJECTIVES

I first used multiple logistic regression to study the relation between flavonoids and cancer risk. In particular, I investigated flavonoids in relation to the risk of various neoplasms in a series of case-control studies conducted in Italy [21] and a study conducted in Greece [22], which included cancers of the oral cavity and pharynx, esophagus, larynx, stomach, colorectum, liver, pancreas, breast, ovary, prostate, and kidney.

I then developed a standardized method to compute the three measures of TAC (TEAC, TRAP and FRAP) in order to take into account the total amount of antioxidants deriving by diet in the analysis for flavonoids and to investigate TAC in relation to the risk of various cancers.

I focused my research on dealing with the high collinearity problems the I encountered in the analysis on colorectal and oral and pharyngeal cancers. In particular, the following issues have been addressed by using different statistical techniques of multivariate analysis:

1 collinearity within classes of flavonoids: which class of flavonoids is associated with cancer risk adjusting for other classes of flavonoids?

2 collinearity between flavonoids and some major food sources of flavonoids associated to cancer risk, especially wine: how to estimate the association between flavonoids and cancer risk excluding the confounding of detrimental effect of wine consumption?

3 collinearity between flavonoids and other biocative compounds contained in plant foods (vitamins, carotenoids, etc.): how to estimate the association between flavonoids and cancer risk excluding the confounding effect of compounds deriving by food sources similar to those of flavonoids?

At last, I took into account the bias due to the possible over-reporting of cases in case-control studies by proposing a new statistical approach, based on the residual method.

3. MATERIAL AND METHODS

3.1 Description of data

I used data from a series of case-control studies conducted in various regions of Italy since the early 1990s [21]. These studies included a total of 9622 cancer cases and 16 050 controls. In particular, the study on oral cavity and pharynx cancers was conducted between 1992 and 2005 in Northern and Central Italy and included 805 cases and 2081 controls; the study on esophageal cancer was conducted between 1992 and 1997 in Northern and Central Italy and included 304 cases and 743 controls; the study on laryngeal cancer was conducted between 1992 and 2000 in Northern Italy and included 460 cases and 1088 controls; the study on stomach cancer was conducted between 1997 and 2007 in the Greater Milan area and included 230 cases and 547 controls; the study on colorectal cancer was conducted between 1992 and 1996 in six Italian areas and included 1953 cases and 4154 controls; the study on pancreatic cancer was conducted between 1991 and 2008 in northern Italy, and included 326 cases and 652 frequency matched control; the study on breast cancer was conducted between 1991 and 1994 in six Italian areas and included 2569 cases and 2588 controls; the study on ovarian cancer was conducted between 1992 and 1999 in Northern and Central Italy and included 1031 cases and 2411 controls; the study on prostatic cancer was conducted between 1991 and 2002 in four Italian areas and included 1294 cases and 1451 controls; the study on renal cancer was conducted between 1992 and 2004 in four Italian areas and included 7670 cases and 1534 controls.

I also analyzed data from a case-control study on liver cancer conducted in Greece between 1995 and 1998 in three teaching hospitals of Athens; the study included 230 cases and 547 controls.

In Italian studies, cases were individuals admitted to hospitals with incident, histologically confirmed cancer, and controls were patients with no history of cancer admitted to the same hospitals for acute, nonneoplastic conditions. Centrally trained interviewers administered a standard

questionnaire to cases and controls during their hospital stay. The questionnaire included personal and socio-demographic characteristics, anthropometric measures, and lifestyle habits, including tobacco smoking and alcohol consumption. A reproducible [23] and valid [24] food frequency questionnaire (FFQ) was used to assess the patients' usual diet in the 2 years preceding diagnosis (for cases) or hospital admission (for controls). The FFQ included the average weekly consumption of 78 food items or food groups and beverages. Intakes lower than once a week, but at least once per month were coded as 0.5 per week.

In the Greek study, cases were incident and histologically confirmed cancers and controls were patients hospitalized with non-cancer disorders usually requiring some minor operations. All subjects were interviewed in the hospital using a standardized semiquantitative FFQ including 100 different foods or beverages, which has been subsequently validated [25, 26]. Data concerning demographic, socio-economic, and medical variables were recorded, and detailed histories smoking habits and 80 alcohol consumption were taken.

3.2 Dietary factors estimates

I developed a standardized method based on food composition tables in terms of flavonoids and proanthocyanidins in order to translate the frequency of consumption of each food item of the FFQ into average daily intakes of flavonoids and proanthocyanidins, taking into account the portion size of each item food. For the six major classes of flavonoids, i.e., flavanols, flavanones, flavonols, anthocyanidins, flavones, and isoflavones, I used food composition data published by the US Department of Agriculture (USDA) [10, 11], further integrated with other sources when needed [27-29]. Major flavonoids included in the six classes were epicathechin and catechin for flavanols, hesperetin and naringerin for flavanones, quercetin for flavonols, cyanidin and malvidin for anthocyanidins, apigenin and luteolin for flavones, and daidzein and genistein for isoflavones. In our control population, flavanols came mainly from tea, apples, pears and wine; flavanones from

oranges and other citrus fruits; flavonols from apples, pears and various common vegetables; anthocyanidins from wine, strawberries, cherries and onions; flavones from cooked vegetables and tea; and isoflavones from soya and bean soups (Table 1).

For proanthocyanidins, data from the USDA became available more recently [12]. Since analytical technology did not allow quantification of these compounds according to their type linkage (e.g., procyanidins, prodelphinidins, etc.), but only according to their degree of polymerization, the USDA food composition tables were in terms of six classes of proanthocyanidins, i.e., monomers, dimers, trimers, 4-6 mers, 7-10 mers, > 10 mers [12]. Given the high correlation between some classes of proanthocyanidins, I further combined monomers and dimers, as well as polymers with three or more mers, and also studied total proanthocyanidins. In our data, the major sources of combined monomers and dimers of proanthocyanidins were wine, apples or pears, peaches, apricots or prunes, whereas major sources of proanthocyanidins with three or more mers were apples or pears, wine, vegetables or bean soup (Table 1). Other main sources are chocolate, pulses and grape.

Nutrients and energy intakes were computed using an Italian food composition database, integrated with others published data [30, 31]. Greek food composition tables for the liver cancer study were used [32].

Similarly, I quantified the TAC of the diet of each subject by computing TEAC (Trolox Equivalent Antioxidant Capacity), TRAP (Total Radical-trapping Antioxidant Parameter), and FRAP (Ferric Reducing-Antioxidant Power) through Italian food tables in terms of these three assays recently published from the National Institute for Food and Nutrition [33]. Concerning TEAC, the assay measures the ability of antioxidant molecules to quench the long-lived $ABTS^{\cdot+}$, compared with that of Trolox, a water-soluble vitamin E analog. TEAC is expressed in mmol of Trolox per kg (solid foods and oils) or per L (beverages) of sample. For TRAP, the assay gives a measure of the protection provided by antioxidants on the fluorescence decay of R-phycoerythrin (lag-phase) during a controlled peroxidation reaction. TRAP values were calculated from the length of the lag-

phase due to the sample compared with that of Trolox and expressed as mmol of Trolox per kg (solid foods) or per L (beverages) of sample. For FRAP, the assay measures the reduction of the $Fe3^{+}$-TPTZ complex to the ferrous form at low pH. FRAP values were obtained by comparing the absorption change in the test mixture with those obtained from increasing concentrations of $Fe3^{+}$ and expressed as mmol of $Fe2^{+}$ equivalents per kg (solid food) or per L (beverages) of sample. The principal food sources of TAC included wine, citrus fruits, apples and pears, and bread.

3.3 Statistical analysis

3.3.1 Logistic regression

I derived the ORs and the corresponding 95% confidence intervals (CIs), by multiple logistic regression models [34] including terms for study centre, sex, age, education, energy intake, as well as other major recognized confounding factors for each cancer of interest. These included alcohol consumption, tobacco smoking and body mass index (BMI) for upper aerodigestive tract neoplasms; calendar year of interview, tobacco smoking, and BMI for stomach cancer; alcohol consumption, BMI, occupational physical activity, family history of colorectal cancer for colorectal cancer; year of interview, alcohol consumption, tobacco smoking, history of diabetes for pancreatic cancer; alcohol consumption and parity for breast cancer; alcohol consumption, parity, oral contraceptives and family history of ovarian and/or breast cancer for ovarian cancer; BMI and family history of prostate cancer for prostate cancer; alcohol consumption, tobacco smoking, BMI, occupational physical activity, and family history of kidney cancer for renal cell carcinomas. According to the study design, I used unconditional or conditional logistic models, matched for study centre, sex, and age. Terms for age were generally entered in the models as quinquennia (categorically), education as three categories (<7, 7-11, ≥ 12 years, categorically), alcohol consumption as quartiles (categorically) and number of drinks (continuously), tobacco smoking as four categories (ex smokers, <25, ≥ 25 cigarettes per day, categorically), BMI as quintiles

(categorically), physical activity as three categories (low, medium, and high, categorically), family history of cancer as dummy (Yes/No).

Adjustment for energy intake was performed entering the term in the model (with or without energy from alcohol intake) and using the residual method [35]. Because both analyses yielded similar results, only the former estimates generally were presented.

In the Greek liver cancer study, odds ratios and 95% CI 108 were estimated by modelling the data through logistic regression, including terms for age (quinquennia, categorically), years of education (<12, ≥12 years), tobacco smoking (never smokers, ever smokers of < 25, ≥ 25 cigarettes per day), and total energy intake (quintiles, categorically).

Flavonoids or the variables of interest were entered in the models as quintiles or tertiles computed on the distribution of controls. ORs per an increment of intake equal to one standard deviation based on control distribution were also computed. Tests for trend for quintiles were based on the likelihood ratio test between the models with and without a linear term for quintiles.

I separated the relation with flavonoids and proanthocyanidins from their major sources by mutual adjustment. Similarly, I computed ORs for flavonoids and proanthocyanidins mutually adjusted for other classes of flavonoids and proanthocyanidins. I also examined additional models including terms for intakes of vitamin C, vitamin E, potassium and folate. In case of a high collinearity, I compared the standardized regression coefficients.

Moreover, I estimated the ORs for flavonoid and proanthocyanidin intake in the models further adjusted for TAC intake in order to investigate whether antioxidant capacity could explain the effect of some classes of flavonoids and proanthocyanidins on colorectal cancer risk or the associations with flavonoids and proanthocyanidins were independent of TAC.

3.3.2 Factor analysis

I used factor analysis (FA) in order to describe the “Fconstruct of flavonoids “ in terms of a minor underlying quantities (called *factors*) [36]. If flavonoids within a particular group are highly

correlated among themselves and have relatively small correlations with variables in a different group, then each group of variables may represent a single underlying factor.

Among the assumption of exploratory FA there are large sample size, continuous distributions, and linear relationship among variables, typical for Pearson correlation.

Let X be the observable random vector with *p* components, mean μ and covariance matrix Σ. The factor model postulates that X is linearly dependent upon a few unobservable random variables F_1, F_2, …, F_m, called *common factors*, and *p* additional sources of variations ε_1, ε_2, …, ε_p, called *errors*. In particular, the FA model is:

$$\begin{aligned} X_1 - \mu_1 &= l_{11}F_1 + l_{12}F_2 + \ldots + l_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= l_{21}F_1 + l_{22}F_2 + \ldots + l_{2m}F_m + \varepsilon_2 \\ &\vdots \\ X_p - \mu_p &= l_{p1}F_1 + l_{p2}F_2 + \ldots + l_{pm}F_m + \varepsilon_p \end{aligned}$$

or, in matrix notation:

$$\underset{(p\text{x}1)}{\text{X} - \mu} = \underset{=(p\text{x}m)}{\text{L}} \underset{(m\text{x}1)}{\text{F}} + \underset{(p\text{x}1)}{\varepsilon}$$

The coefficient l_{ij} is called *factor loading* of the *i*th variable on the *j*th factor, so the matrix L is the *matrix of factor loadings*. The *i*th specific factor ε_i is associated only with variable X_i. The *p* deviations $X_1 - \mu_1$, $X_2 - \mu_2$, ..., $X_p - \mu_p$ are expressed in terms of $p + m$ random variables F_1, F_2, ..., F_m, ε_1, ε_2, ..., ε_p which are unobservable.

Moreover it is assumed that:

$$\underset{(m\text{x}1)}{E(\text{F})} = 0, \quad \text{Cov(F)} = \underset{(m\text{x}m)}{E[\text{FF}']} = \text{I}$$

$$\underset{(p\text{x}1)}{E(\varepsilon)} = 0,\ \text{Cov}(\varepsilon) = \underset{(pxp)}{E[\varepsilon\varepsilon']} = \Psi = \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{pmatrix}$$

and that F and ε are independent, so:

$$\text{Cov}(\varepsilon, F) = \underset{(pxm)}{E(\varepsilon F')} = 0$$

The model with these assumptions constitutes the *orthogonal factor model*. If the matrix Cov(F) was not diagonal and consequently the factors were not independent but correlated, we would obtain the *oblique model* that presents additional estimation difficulties.

The *orthogonal factor model* implies the following covariance structure for X:

$$(X - \mu)(X - \mu)' = (LF + \varepsilon)(LF + \varepsilon)$$

$$= LF(LF)' + \varepsilon(LF)' + LF\varepsilon' + \varepsilon\varepsilon'$$

so that

$$\Sigma = \text{Cov}(X) = E(X - \mu)(X - \mu)'$$

$$= LE(FF')L' + E(\varepsilon F')L' + LE(F\varepsilon') + E(\varepsilon\varepsilon')$$

$$= LL' + \Psi \qquad (*)$$

given that according to the previous assumptions Cov(F)=0, Cov(ε), and Cov(ε,F) = $E(\varepsilon,F')$= 0.

Moreover, $(X - \mu)F' = (LF + \varepsilon)F' = LFF' + \varepsilon F'$, so

$$\text{Cov}(X, F) = E(X - \mu)F' = LE(FF') + E(\varepsilon F') =$$

$$= L \qquad (**)$$

given that Cov(F)=0 and $E(\varepsilon,F')$= 0.

The equations (*) and (**) imply that

$$Var(X_i) = l_{i1}^2 + l_{i2}^2 + ... + l_{im}^2 + \psi_i$$
$$Cov(X_i, X_k) = l_{i1}l_{k1} + ... + l_{im}l_{km}$$
$$Cov(X_i, F_j) = l_{ij}$$

The model $X - \mu = L\ F + \varepsilon$ is linear in the common factors, and the assumption of linearity is inherent in the formulation of factor model. In fact, if the relation is not linear the covariance structure may not be adequate (e.g., if $x_1 - \mu_1 = l_{11}F_1F_3 + \varepsilon_1$ and $x_2 - \mu_1 = l_{21}F_2F_3 + \varepsilon_2$).

The last equations are useful to see the $\mathrm{Var}(X_i) = \sigma_{ii}$ as a portion of variance of the *i*th variable contributed by the *m* common factors, called *i*th *communality, and a* portion of the specific *i*th factor, called the *uniqueness*, or *specific variance:*

$$\sigma_{ii} = \underbrace{l_{i1}^2 + l_{i2}^2 + \ldots + l_{im}^2}_{\text{communality}} + \psi_i$$

$$\mathrm{Var}(X_i) = \text{communality} + \text{specific variance}$$

Indicating the communality with h_i^2, $h_i^2 = l_{i1}^2 + l_{i2}^2 + \ldots + l_{im}^2$ we obtained:

$$\sigma_{ii} = h_i^2 + \psi_i \quad i = 1, 2, \ldots, p$$

where the *i*th communality is the sum of squares of the loadings of the *i*th variable on the *m* common factors.

The factor model assumes that the $p + p(p-1)/2 = p(p+1)/2$ variances and covariances for X can be reproduced from the *pm* factor loadings l_{ij} and the *p* specific variances ψ_i. When m=p any covariance matrix can be reproduced exactly, but is when *p* is small relative a *m* that factor analysis is most useful because the factor model can provide a simple explanation of the covariation in X with fewer parameters than the *p(p+1)*/2 parameters in Σ (e.g., if p=12 and want to describe the structure with 2=*m* factors, we will be able to describe the entire covariance with 36 parameters instead of 78).

When $m > 1$, there is not a unique solution to the equation $\Sigma = LL' + \Psi$. In fact, let T be any *m* x *m* orthogonal matrix so that $TT' = T'T = I$, the factor model can be written as:

$$X - \mu = LF + \varepsilon = LTT'F + \varepsilon = L^*F^* + \varepsilon$$

where

$$L^* = LT \text{ and } F^* = T'F$$

Since

$$E(F^*) = T'E(F) = 0$$

and

$$\text{Cov}(F^*) = T'\text{Cov}(F)T = T'T = I_{(m \times m)}$$

It is impossible on the basis of observation on X, to distinguish the loadings L from the loadings L^*. That is, the factors F and $F^* = T'F$ have the same statistical proprieties, and even if the loadings L^* are different from the loadings L, they both generate the same covariance matrix, that is

$$\Sigma = LL' + \Psi = LTT'L + \Psi = (L^*)(L^*)' + \Psi$$

This ambiguity provides the rationale for the *factor rotation*, since orthogonal matrices correspond to rotations (and reflections) of the coordinate system for X.

Fcator loading L are determined only up to an orthogonal matrix T. This, the loadings L*=LT and L both give the same representation. The communalities, given by the diagonal elements of LL'=L*L*' that is $l_{i1}^2 + l_{i2}^2 + ... + l_{im}^2$ are not affected by the choice of F. This means that given *m*, for each choice of T and consequently F and L, the variance of Xi explained by *m* factors is always the same ($l_{i1}^2 + l_{i2}^2 + ... + l_{im}^2$) for each i.

We do impose conditions if we want to uniquely estimate L and Ψ. So we generally fix the rotation (T) according to a criterion that facilitate the interpretation of factors. Once the loadings and specific variances (ψ_i) are obtained, factors are estimated.

Model estimation

Let S be the sample covariance matrix of *n* observations $x_1, x_2, \ldots, x_n$ on *p* variables. It is an estimator of the unknown population covariance matrix Σ. If the off-diagonal elements of S are small or those of the sample correlation matrix R are essentially zero, the variables are not related and thus a FA should not be performed. In such a situation, the *uniqueness* play a dominant role, while the major purpose of FA is to determine a few important *common factors*, that adequately describe the phenomenon under consideration.

If Σ appears to deviate significantly from a diagonal matrix, then a FA can be applied, and the initial problem is to estimate the factor loadings l_{ij} and the specific variances ψ_i.

Among the methods of parameter estimation, the most common are the *principal component method*. Let $(\hat{\lambda}_1, \hat{\mathbf{e}}_1)$, $(\hat{\lambda}_2, \hat{\mathbf{e}}_2)$, …, $(\hat{\lambda}_\mathrm{p}, \hat{\mathbf{e}}_\mathrm{p})$, with $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_\mathrm{p}$. be the eigenvalue-eigenvector pairs of the matrix S. Since this matrix is symmetric, it can be written as the following spectral decomposition

$$S = \hat{\lambda}_1 e_1 e_1' + \hat{\lambda}_2 e_2 e_2' + \ldots + \hat{\lambda}_p e_p e_p'$$

$$= [\sqrt{\hat{\lambda}_1} e_1 \,|\, \sqrt{\hat{\lambda}_2} e_2 \,|\, \ldots \,|\, \sqrt{\hat{\lambda}_p} e_p] \begin{bmatrix} \sqrt{\hat{\lambda}_1} e_1' \\ \sqrt{\hat{\lambda}_2} e_2 \\ \ldots \\ \sqrt{\hat{\lambda}_p} e_p' \end{bmatrix}$$

Since we are interested to describe S in terms of $m<p$ common factors, S can be approximated as follows

$$\Sigma \doteq [\sqrt{\hat{\lambda}_1} e_1 \,|\, \sqrt{\hat{\lambda}_2} e_2 \,|\, \ldots \,|\, \sqrt{\hat{\lambda}_m} e_m] \begin{bmatrix} \sqrt{\hat{\lambda}_1} e_1' \\ \sqrt{\hat{\lambda}_2} e_2' \\ \ldots \\ \sqrt{\hat{\lambda}_m} e_m' \end{bmatrix} + \begin{bmatrix} \hat{\psi}_1 & 0 & \cdots & 0 \\ 0 & \hat{\psi}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\psi}_p \end{bmatrix}$$

$$= \hat{L}\hat{L}' + \hat{\Psi}$$

With $\tilde{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_1 \,\middle|\, \sqrt{\hat{\lambda}_2}\hat{\mathbf{e}}_2 \,\middle|\, \ldots \,\middle|\, \sqrt{\hat{\lambda}_\mathrm{m}}\hat{\mathbf{e}}_\mathrm{m} \right]$ the matrix of the estimated factor loadings.

The estimated specific variances are provided by the diagonal elements of the matrix $S - \hat{L}\hat{L}'$, so

$$\tilde{\Psi} = \begin{bmatrix} \tilde{\psi}_1 & 0 & \cdots & 0 \\ 0 & \tilde{\psi}_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{\psi}_1 \end{bmatrix} \quad \text{with } \tilde{\psi}_i = s_{ii} - \sum_{j=1}^{m} \tilde{l}_{ij}^{\,2}$$

Communalities are estimated as

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + ... + \hat{l}_{im}^2$$

The solutions of these methods can be rotated in order to simplify the interpretation of factors.

Note that in cases in which the units of the variables are not commensurate, it is usually desirable to work with the standardized variables

$$z_j = \begin{bmatrix} \frac{(x_{j1} - \bar{x}_1)}{\sqrt{s_{11}}} \\ \frac{(x_{j2} - \bar{x}_2)}{\sqrt{s_{22}}} \\ ... \\ \frac{(x_{jp} - \bar{x}_p)}{\sqrt{s_{pp}}} \end{bmatrix} \quad j=1,2,...,n$$

whose sample covariance matrix is the sample correlation matrix R of the observations $x_1, x_2, \ldots, x_n$. The principal component FA of the sample correlation matrix is obtained starting with R in place of S.

Choice of the number of factors

If the number of common factors is not determined by a priori considerations, the choice of *m* can be based on the estimated eigenvalues. In fact, it is possible to show that the sum of squared entries of the *residual matrix* $S - \hat{L}\hat{L}' + \tilde{\Psi}$ is $\leq \hat{\lambda}^2_{m+1} + ... + \hat{\lambda}^2_p$, and this implies that a small value for the sum of the squares of the eigenvalues gives a small value for the sum of squared errors of approximation. The contribution to the total sample variance from the first common factor is $\hat{l}_{11}^2 + \hat{l}_{21}^2 + ... + \hat{l}_{p1}^2 = \left(\sqrt{\hat{\lambda}_1}\hat{e}_1\right)'\left(\sqrt{\hat{\lambda}_1}\hat{e}_1\right) = \lambda_1$ since $\hat{e}_1$ has unit length. Proportion of total sample variance due to *j*-th factor is equal to $\frac{\hat{\lambda}_j}{p}$ if the variables are standardized.

The choice of the number of factors was based on three main criteria. The first one is to retain those factors with eigenvalues greater than 1.00. The second criterion is to add successive factors until the cumulative percentage of variance explained by the retained factors is satisfactory. To

terminate the factor extraction process, we considered 75-80% to be a valid threshold for the cumulative variance extracted. The third one, suggested by Cattell [37], is to plot, by the option SCREE in SAS, the extracted factors against their eigenvalues in descending order of magnitude to identify distinct breaks in the slope of the plot, called "scree plot". To determine where the break occurs, a straight line should be drawn with a ruler through the lower values of the plotted eigenvalues. That point where the factors curve above the straight line drawn through the smaller eigenvalues identifies the optimal number of factors to retain.

Finally, to determine the number of factors to retain, a researcher should not be based only on statistical criteria, but also on subjective motivations. In fact, the other criterion to take in account is factor interpretability.

Factor rotation

All factor loadings obtained from the initial loadings by an orthogonal transformation have the same ability to reproduce the covariance (or correlation) matrix. If $\hat{\mathbf{L}}$ is the *p* x *m* matrix of estimated

factor loadings obtained by any method, then

$$\hat{L}^* = \hat{L}T \quad \text{where } TT' = T'T = I$$

is a *p* x *m* matrix of "rotated" loadings. The estimated covariance (or correlation) matrix remains unchanged, since:

$$\hat{L}\hat{L}'+\hat{\Psi} = \hat{L}TT'\hat{L}'+\hat{\Psi} = \hat{L}^*\hat{L}^{*\prime}+\hat{\Psi}$$

This equation indicates that also the residual matrix remains unchanged:

$$S_n - \hat{L}\hat{L}'-\hat{\Psi} = S_n - \hat{L}^*\hat{L}^{*\prime}-\hat{\Psi}$$

Moreover the specific variances $\hat{\psi}_i$, and so the communalities $\hat{h}_i^2$, are unchanged as well. Thus, from a mathematical point of view, it doesn't make any difference whether $\hat{L}$ or $\hat{L}^*$ is obtained. Since the original loadings may not be easily interpretable, it is usual practice to rotate them until a "simple structure" is obtained. Ideally, it is desirable to have a pattern of loadings such that each

variable loads highly on a single factor and has small-to-moderate loadings on the remaining factors. This solution is computed by selecting the orthogonal transformation T that maximizes the variance V of factor loadings, that is to find $\tilde{l}$ such that

$$V=\frac{1}{p}\sum_{j=1}^{m}\left[\frac{\sum_{j=1}^{p}\left[l_{ij}^{4}\right]-\left(\sum l_{ij}^{2}\right)^{2}}{p}\right]$$

is maximized.

This simplest case of rotation is an *orthogonal rotation* in which the angle between the reference axes of factors are maintained at 90 degrees; this implies that the rotated factors remain uncorrelated. Other forms of rotation, indicated as *oblique rotations*, allow the angle between the reference axes to vary, i.e., factors are allowed to be correlated with each other. Orthogonal rotation procedures are more commonly used than oblique rotation procedures, and should be performed when the common factors are assumed to be independent. This is a crucial aspect in nutritional epidemiology, where one may deal with severe multicollinearity problems, and so I used forms of *orthogonal* rotation. In other situations where the correlations between the underlying constructs are not assumed to be zero, oblique rotations may yield simpler and more interpretable factor solutions.

Measures of sampling adequacy

In order to test whether the sample covariance matrix is factorable, I used measures of sampling adequacy that compare the simple and partial correlation coefficients either overall or for single variables. The overall measure, called Kaiser-Meyer-Olkin statistic (*KMO*), is defined as follows [38]:

$$KMO=\frac{\sum_{i\neq j}\sum r_{ij}^{2}}{\sum_{i\neq j}\sum r_{ij}^{2}+\sum_{i\neq j}\sum a_{ij}^{2}}$$

where $\sum\sum$ are the sum over all variables in the matrix when variable $i \neq$ variable j, r_{ij} is the Pearson correlation coefficient between i and j, and a_{ij} the partial correlation coefficient between i and j. Individual measures of sampling adequacy are computed using only the simple and partial correlation coefficients involving the specific variable under consideration. The overall and individual measures range between 0 and 1. Smaller values indicate that the squared correlation coefficient is small relative to the squared partial correlation coefficient and therefore a FA may be imprudent. If the sum of the squared partial correlation coefficients is small compared with the sum of the squared correlation coefficients, the measures approach 1.

On the contrary, if the off-diagonal elements sample correlation matrix are very small, the variables are not related and thus a FA should not be performed. To test such a situation, we used Bartlett's test of sphericity that tests the null hypothesis that the correlation matrix is an identity matrix. It is a chi-square test [38], whose statistic is defined as follows:

$$\chi^2 = -\left[(N-1) - \left(\frac{2k+5}{6}\right)\right] \log|R|$$

where χ^2 is the calculated chi-square value for Bartlett's test, N is sample size, k is the number of variables in the matrix and $|R|$ the determinant of the correlation matrix. The degrees of freedom for this chi-square are $k(k-1)/2$. Larger values of the test suggest that the null hypothesis should be rejected.

4. RESULTS - FLAVONOIDS AND CANCER RISK

4.1 Cancers at the upper aerodigestive tract

4.1.1 Flavonoids and oral and pharyngeal cancer

Data on oral and pharyngeal cancer (Figure 1 – Appendix 1) showed that total flavonoids were inversely related to the risk of this neoplasm [39]. The ORs for the highest versus the lowest quintile of all classes of flavonoid intake were below unity. The OR was 0.51 for flavanones, 0.62 for flavonols, and 0.56 for total flavonoids, with significant inverse trend in risk. No significant association emerged for other classes of flavonoids. The ORs were consistent across strata of age, sex, education, BMI, tobacco smoking and alcohol intake. After allowance for vegetable and fruit consumption, the inverse relations with total flavonoids and flavanones remained significant, whereas that with flavonols became non significant. None of the associations was significant after further allowance for vitamin C, probably on account of the high correlation between these compounds.

4.1.2 Flavonoids and esophageal cancers

Flavanones were significantly inversely associated with esophageal cancer risk (OR, 0.38) (Figure 1 – Appendix 1) [40]. The inverse relation tended to be stronger in subjects who drank ≥ 6 drinks of alcoholic beverages per day. After allowance for fruit intake or vitamin C, the association of flavanones with esophageal cancer remained inverse, though non-significant, suggesting that flavanone may explain, together with vitamin C, the protective effect of citrus fruits on esophageal cancer, since citrus fruit accounts for 90% of flavanone intake.

4.1.3 *Flavonoids and laryngeal cancers*

Significant inverse relations were found between flavanols (OR,0.64), flavanones (OR, 0.60), flavonols (OR, 0.32), and total flavonoids (OR, 0.60) and laryngeal cancer risk (Figure 1 –

Appendix 1), although the trends in risk were significant only for flavanones and flavonols [41]. The estimates changed little after controlling for vegetable, fruit, and vitamin C intake.

4.2 Cancers at the digestive tract

4.2.1 Flavonoids, proanthocyanidins and stomach cancer

Although all ORs were below unity, no significant association emerged for any class of flavonoids (Figure 2 – Appendix 1) [42]. There was an inverse trend in risk for flavonols, but the continuous OR (0.88) was not significant. There were inverse associations between all classes of proanthocyanidins and stomach cancer risk, with significant trend in risk: the ORs were 0.44 for monomers and dimers combined, 0.36 for polymers with three or more mers, and 0.34 for total PAs. These estimates did not change when I adjusted for total flavonoid intake, and for fruit and vegetable consumption.

4.2.2 Flavonoids, proanthocyanidins and colorectal cancer

Significant inverse trend in the risk of colorectal cancer was found with increasing intake of anthocyanidins (OR, 0.67), flavonols (OR, 0.64), flavones (OR, 0.78), and isoflavones (OR, 0.76) (Figure 3 – Appendix 1) [43]. The estimates did not substantially differ for colon and rectal cancers. After allowance for fruit and vegetables consumption, dietary fiber, or certain micronutrients including vitamin C, the associations with flavonoids did not change by more than 10%.

A significant trend of decreasing colorectal cancer risk emerged also with increasing intake of proanthocyanidins, for all classes, except for monomers [44]. The OR for the highest versus the lowest quintile of intake was 0.82 for monomers and dimers combined, and 0.74 for all polymers with three or more mers (0.84 for trimers, 0.80 for 4-6 mers, 0.79 for 7-10 mers, 0.69 for polymers more than 10 mers). Adjustments for other classes of flavonoids did not substantially modify the associations between proanthocyanidins and colorectal cancer risk. The inverse associations appeared to be stronger for rectal than for colon cancer.

4.2.3 Flavonoids and liver cancer

Table 1 (Appendix 1) shows multiple ORs and corresponding 95% CI of six classes of flavonoids among hepatitis B (HBV) and/or C (HCV) virus postive, HBV and HCV negative HCC cases, and cholangiocarcinoma (CAC) cases [15]. There were no distinct patterns with respect to either HCC virus positive or HCC virus negative in relation to total flavonoids or any class of flavonoids, with the exception of flavones. For the latter class, there were inverse associations with respect to both HCC virus positive category (P-trend, 0.049) and HCC virus negative category (P-trend, 0.084). If we ignore the distinction into virus positive or negative and evaluate HCC cases as a total, the inverse association with flavones was more evident and statistically significant (P-trend, 0.023). With respect to the generally rare CAC, there were apparently strong inverse associations with flavan-3-ols and anthocyanidins, as well as total flavonoids.

4.2.4 Flavonoids, proanthocyanidins and pancreatic cancer

Inverse associations were found with significant trend in risk for isoflavones (OR, 0.40) and flavanones (OR, 0.68), and with a borderline significant trend in risk for flavonols (OR, 0.69) (Figure 4 - Appendix 1) [45]. No meaningful association emerged for other flavonoids, including anthocyanidins (OR, 0.83), flavanols (OR, 0.63), and flavones (OR, 0.88). All the ORs were below unity for all proanthocyanidins with a significant trend in risk for polymers with three or more mers and total proanthocyanidins. The ORs were similar in all classes of polymers with three or more mers and in their combinations (OR, 0.41). After adjustment for fruit and vegetables, the OR for flavonoids and proanthocyanidins did not substantially change. After adjustment for vitamin C or folate, the associations for isoflavones and proanthocyanidins weakened, whereas the associations with other classes of flavonoids disappeared.

4.3 Hormone-related cancers

4.3.1 Flavonoids and breast cancer

A reduced risk of breast cancer was found for increasing intake of flavones (OR, 0.81, p-trend=0.02), and flavonols (OR, 0.80, p-trend, 0.06) (Figure 5 – Appendix 1) [46].

4.3.2 Flavonoids and ovarian cancer

Flavonols (OR, 0.63) and isoflavones (OR, 0.51) were inversely related to ovarian cancer with a significant trend in risk (Figure 5 – Appendix 1) [47]. Further adjustments for fruit and vegetable intake did not materially modify these associations.

4.3.3 Flavonoids and prostatic cancer

No association was found between any of the analyzed class of flavonoids and prostate cancer risk [48].

4.4 Other cancers

4.4.1 Flavonoids and renal cancer

Flavonols (OR, 0.69) and flavones (OR, 0.68) were significantly inversely related to the risk of renal cancer [49] (Figure 5 – Appendix 1). Allowance for vegetable and fruit consumption only partly explained the inverse relation with these flavonoids.

5. RESULTS - TAC AND CANCER RISK

5.1 TAC, flavonoids, proanthocyanidins and colorectal cancer

Table 1 (Appendix 2) shows the mean daily intake and the upper cut off points of quintiles of TAC intakes among controls. The mean daily intake was 4.47 mmol for TEAC, 4.56 mmol for TRAP, 11.45 mmol for FRAP.

The three indexes were strongly correlated: the correlation was 0.97 between TEAC and TRAP, 0.99 between TEAC and FRAP, and 0.98 between TRAP and FRAP. For this reason, I would have expected to obtain similar results; since I analyzed TAC for the first time, I present results on the three indexes.

Table 2 (Appendix 2) gives the correlations of TEAC, TRAP and FRAP with selected covariates including fruit, vegetable, flavonoids and PAs, among controls. High correlation was found between the three TAC indexes and flavonoids (~0.60), anthocyanidins (~0.85) and proanthocyanidins (~0.60).

Table 3 (Appendix 2) gives the ORs of colorectal cancer according to quintiles of TAC. We found inverse associations between all three measures of TAC and colorectal cancer risk, with OR estimates very similar among the three measure of TAC, as expected. The OR for the highest versus the first quintile was 0.78 (95% CI, 0.63-0.96; p-trend, 0.002) for TEAC, 0.74 (95% CI, 0.60-0.92; p-trend, 0.001) for TRAP and 0.78 (95% CI, 0.63-0.96; p-trend, 0.003) for FRAP. ORs did not materially change after adjustment for fruit and vegetables. When I separately studied the cancer of rectum and of colon, I found that the OR was 0.89 (95% CI, 0.70-1.15; p-trend, 0.15) for TEAC, 0.82 (95% CI, 0.63-1.05; p-trend, 0.037) for TRAP and 0.85 (95% CI, 0.66-1.09; p-trend, 0.13) for FRAP for colon cancer; and 0.65 (95% CI, 0.48-0.88; p-trend, 0.001) for TEAC, 0.66 (95% CI, 0.49-0.90; p-trend, 0.002) for TRAP and 0.70 (95% CI, 0.52-0.95; p-trend, 0.002) for FRAP for rectal cancer.

The OR for FRAP became 1.09 (95% CI, 0.82-1.46) adjusting for anthocyanidins, 0.79 (95% CI, 0.64-0.98) adjusting for flavones, 0.90 (95% CI, 0.72-1.13) adjusting for flavonols, 0.88 (95% CI, 0.69-1.11) adjusting for PAs, 0.87 (95% CI, 0.69-1.08) adjusting for proanthocyanidins with more than 10 mers. Similar trend was found in the estimates of ORs for TEAC and TRAP.

Conversely, after adjusting for FRAP, the OR was 0.63 for anthocyanidins (95% CI, 0.48-0.84), 0.80 for flavones (95% CI, 0.67-0.95), 0.67 for flavonols (95% CI, 0.55-0.81), 0.81 for total proanthocyanidins (95% CI, 0.65-0.99) and 0.73 for proanthocyanidins with more than 10 mers (95% CI, 0.60-0.89). I obtained similar OR estimates adjusting for TEAC and TRAP rather than FRAP.

6. RESULTS - STATISTICAL TECHNIQUES TO FACE THE PROBLEM OF HIGH COLLINEARITY AMONG DIETARY FACTORS

6.1 Collinearity within flavonoids in colorectal cancer data

The median, mean, standard deviation, maximum and minimum of 34 single classes of flavonoids in 1953 cases and 4154 controls from the study on colorectal cancer are given in Table 1 (Appendix 1). The distributions were generally right-skewed given the presence of outliers.

I applied FA to the 34 single flavonoids, in order to explain the variance of the entire structure of flavonoids with a smaller set of variables, address the problem of collinearity within flavonoids and detect which flavonoids is associated with cancer risk independently from other flavonoids. A different approach consists in grouping the 8 major classes of flavonoids - that is according to their chemical structure - and inserting them in FA instead of the 34 single flavonoids. This is a classical approach since previous studies – including my studies – used these 8 major classes to investigate the relation between flavonoids and cancer risk behind the assumption that flavonoids with similar chemical structure have similar effect on the disease. Although this is an acceptable assumption, it easy to see from Table 1 (Appendix 3) and Table 2 (Appendix 3) that some flavonoids have a very different distribution than other flavonoids belonging to the same class, and since some chemical differences - not in the main structure but in the last ring [4] occurred in these flavonoids, I thought that exploiting the variance of all the 34 flavonoids was much better. For these reason, I started to analyze separately the 34 flavonoids and compared later on these results with those obtained by applying FA to the 8 classes of flavonoids.

6.1.1 Factor analysis on 34 single flavonoids and proanthocyanidins

The ORs for the highest versus the lowest quintile of flavonoid intake were described above and were reported in Figure 1 (Appendix 3). I found inverse association between some classes of flavonoids and colorectal cancer risk, but since the models included only the flavonoids under

investigation and not other flavonoids, the associations were not adjusted for the possible confounding of other flavonoids.

I first standardized the 34 flavonoid variables so that their means were equal to 0 and their variances equal to 1, and then applied FA to these. I found that the correlation matrix was not singular, so there were linear dependencies in the correlation matrix, and some scoring coefficients equal to 0. I first tried excluding proanthocyanidins but I found again a not singular matrix. Possible high correlation values among items could have been attributed to duplication of subjects in the database, but I have preliminary cleaned the data before analysis, and I had enough subject per items. Therefore, in order to verify whether there were too strong correlations among some items, I analyzed the correlations between each couple of flavonoids computed among all subjects (Table 2 – Appendix 3). I found that there were various flavonoids strongly associated with other flavonoids, especially belonging to the same classes, but not only. I started to group flavonoids with correlation coefficients higher than 0.90 and found that they had similar correlation coefficients with all other flavonoids. I first inserted one flavonoid as representative per one group of flavonoids in FA and used the Kaiser-Meyer-Olkin (KMO) test to verify whether the items were still too strongly correlated. In fact, specifying the option *MSA* in the *proc factor*, I obtained the total measure of sampling adequacy (MSA) and the MSA for each item, which were of great help to identify possible errors in grouping flavonoids and improve the choice of flavonoid groups. I did numerous attempts before detecting the groups of flavonoids to insert in the FA by using the same following procedure: 1) I identified the items with high correlation and with the lowest MSA, 2) I removed it from the list of items to be analysed, and 3) I rerun the test of matrices. I then identified the following 15 groups of flavonoids:

1) S1: FL1, ISO-DAIDZEINA ~ FL2, ISO-GENISTEIN
2) S3: FL3, ANTHO-Cyanidin ~ FL6 ANTHO-Pelargonidin

3) S4: FL4, ANTHO-Delphinidin ~ FL5, ANTHO-Malvidin ~ FL7, ANTHO-Peonidin, FL8, ANTHO-Petunidin ~ FL9 FLAVAN3-(+)-Catechin

4) S10: FL 10, FLAVAN3-(-)-Epigallocatechin ~ FL12, FLAVAN3-(-)-Epicatechin 3-gallate ~ FL13, FLAVAN3-(-)-Epigallocatechin 3-gallate ~ FL14, FLAVAN3-Theaflavin ~ FL 15, FLAVAN3-Thearubigins ~ FL25 FLAVAN3-Theaflavin-3,3'-digallate ~ FL26, FLAVAN3-Theaflavin-3'-gallate ~ FL27, FLAVAN3-Theaflavin-3-gallate ~ FL28, FLAVAN3-(+)-Gallocatechin

5) S11: FL11, FLAVAN3-(-)-Epicatechin

6) S16: FL 16, FLAVA-Eriodictyol

7) S17: FL17, FLAVA-Hesperetin ~ FL18, FLAVA-naringenin

8) S19: FL19, FLAVONE-Apigenin

9) S20: FL20, FLAVONE-Luteolin

10) S21: FL21, FLAVONO-Isorhamnetin

11) S22: FL22, FLAVONO-Kaempferol

12) S23: FL23, FLAVONO-Myricetin

13) S24: FL24, FLAVONO-Quercetin

14) S29: FL29, PROANTHO-Monomers ~ FL30, PROANTHO-Dimers

15) S31: FL31, PROANTHO-Trimers ~ FL32, PROANTHO-4-6mers ~ FL33, PROANTHO-7-10mers ~ FL34, PROANTHO-Polymers

In these groups, some classes of flavonoids were divided into various groups and one group was created by flavonoids belonging to different classes, as follows:

Isoflavones: 1 group

Anthocyanidins: 2 groups
Flavan-3-ols: 3 groups
} 1 common group

Flavanones: 2 groups

Flavones: 2 groups

Flavonols: 4 groups

Proanthocyanidins: 2 groups.

This grouping took into account the different distributions of flavonoids belonging to the same class without altering the entire structure of flavonoids that I wanted to explain through FA. Figure 2 and Figure 3 (Appendix 3) show the percentages of intake of each single flavonoid to which various food sources contributed, overall and according to the new grouping. I found that the grouping strongly depended on their major sources. This was not surprising, however, since the variance in flavonoids directly depends on the variance of food consumption from which their intake derives, but it was reassuring for keeping the control of the flavonoid structure.

Once I resolved the problem of multicollinearity, I looked for the items that were not correlated strongly enough with the other items. In fact, since these items have not much shared common variance and would yield as many factors as items in FA, they should be dropped out from FA. I was not able to perform the Bartlett's test of sphericity in SAS in order to test the null hypothesis that the correlation matrix is an identity matrix (i.e. that there is no relationship among the items), and to verify whether exclude or not some factors. However, since the statistic χ^2 is influenced by the sample size, with larger sample sizes resulting in larger values of Bartlett's test, this test would have not been sufficiently informative for studies with large sample size as this study. In fact, it is rare that Bartlett's test is too small to reject the null hypothesis that the matrix should not be factor analyzed, and this test is used only as a minimum standard for assessing the quality of the correlation matrix. I only excluded the group of isoflavones from FA, since their correlation coefficients with all other flavonoids were virtually identical to zero (Table 2 – Annex 3), and remained with sufficient numbers of significant correlations among other items to explain the use of FA.

Thus, applied FA to 14 groups of flavonoids by inserting one flavonoid as representative per one group, but I did not obtain acceptable values for KMO test. When I inserted the sum of flavonoids (instead of a representative flavonoid) per each group, I obtained a total MSA equal to 0.61, a "middling" value according to Kaiser's criteria [38]. Higher values of total MSA (~ 0.67) and of minimum MSA per item (>0.38 versus 0.27) were obtained after exclusion of proanthocyanidins with three or more mers, but I judged that this class of flavonoids was essential in describing the entire structure in an understandable manner.

Once I verified that the correlation matrix was factorable, my aim was to condense the variance that was shared among the items to determine the number of initial subsets or factors that appear to represent the dimensions of the "construct flavonoids". I identified four factors that satisfy the three main criteria used in selecting the number of factors to retain in a FA:

1. *The eigenvalues > 1 criteria*

Eigenvalues of correlation matrix: Total = 14 Mean = 1				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative proportion**
1	5.51275370	3.08424417	0.3938	0.3938
2	2.42850953	0.46184863	0.1735	0.5672
3	1.96666090	0.81124314	0.1405	0.7077
4	1.15541776	0.18323795	0.0825	0.7902
5	0.97217981	0.14972191	0.0694	0.8597
6	0.82245790	0.22934382	0.0587	0.9184
7	0.59311408	0.33708399	0.0424	0.9608
8	0.25603009	0.11014972	0.0183	0.9791
9	0.14588037	0.07932734	0.0104	0.9895
10	0.06655303	0.02914401	0.0048	0.9943
11	0.03740902	0.01383479	0.0027	0.9969
12	0.02357423	0.00769616	0.0017	0.9986
13	0.01587806	0.01229653	0.0011	0.9997
14	0.00358153		0.0003	1.0000

The four eigenvalues grater than 1 and corresponding measures are highlighted in the Table above.

2. *The criteria of cumulative percentages of variance extracted > 75%*

	Factor1	Factor2	Factor3	Factor4
s3	0.31576	0.39937	0.80247	0.01907
s4	0.71535	-0.60604	0.00564	-0.06112
s10	0.31053	0.39036	-0.41660	0.63707
s11	0.85434	-0.27684	0.07383	0.19181
s16	0.18005	0.37859	0.82842	0.14920
s17	0.29569	0.10271	0.30942	0.06323
s19	0.80797	-0.52229	0.02215	0.03270
s20	0.38018	0.49241	-0.22161	0.33862
s21	0.49586	0.51454	-0.08459	-0.62416
s22	0.63625	0.59302	-0.37128	0.07672
s23	0.81469	0.08105	-0.28050	-0.27051
s24	0.80328	0.38657	-0.09857	-0.26705
s29	0.85713	-0.47696	0.03637	0.09202
s31	0.68079	-0.06051	0.27804	0.13514

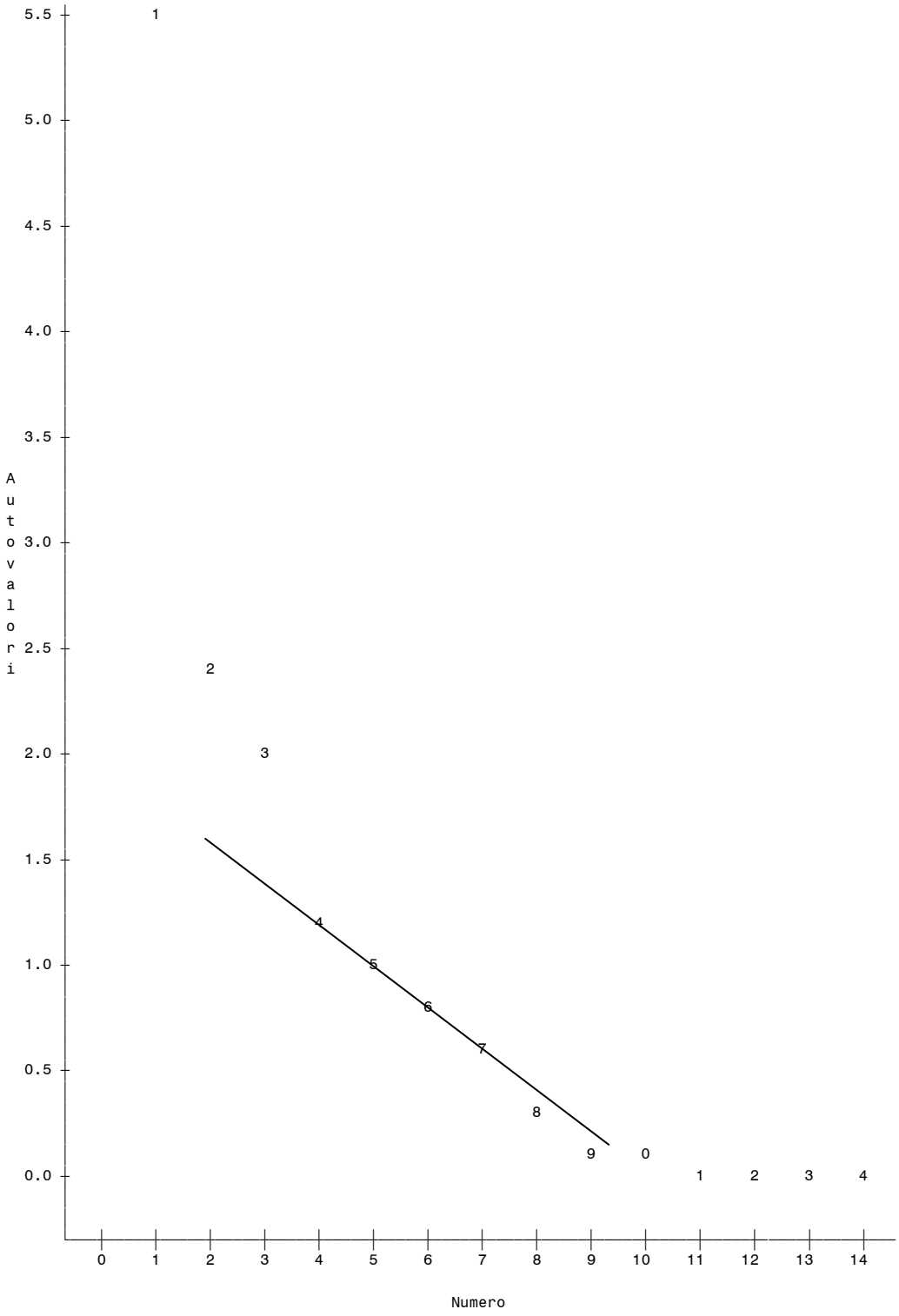
Variance explained by each factor			
Factor1	**Factor2**	**Factor3**	**Factor4**
5.5127537	2.4285095	1.9666609	1.1554178

Item communality (h^2): Total = 11.063342							
s3	**s4**	**s10**	**s11**	**s16**	**s17**	**s19**	**s20**
0.9035240	0.88277088	0.82822383	0.84877181	0.88428782	0.19771488	0.92716240	0.55077815

s21	**s22**	**s23**	**s24**	**s29**	**s31**
0.90735313	0.90021642	0.82213962	0.87572683	0.97195881	0.56271330

The percentage of variance explained was **11.063342/14*100=79%.**

3. The scree plot criteria



Using the Cattel criteria [38], four factors remain above that line that account for the maximum amount of variance in the 14 items.

According to another criteria that factor extraction should be continued until the last factor accounts for only a small portion of the explained variance (less than 5%), I considered also the solution with 6 factors. In fact, the sixth factor had eigenvalue equal to 0.82 and explained 5.9% (=0.82/14*100) of the total variance, whereas the seventh factor had eigenvalue 0.59 and explained 4.2% (0.59/14*100) of the total variance.

In both solutions, the rate of the numbers of factors and the numbers of items remained between ½ and 1/4.

I then applied a Varimax rotation to improve the meaningfulness and interpretation of the generated factors, and obtained the following rotated factor pattern for the solution with four factors.

	Factor1	Factor2	Factor3	Factor4
s3	0.02526	0.15350	0.93762	-0.01398
s4	0.92357	0.10304	-0.05727	-0.12602
s10	0.06221	-0.02233	-0.05727	0.90586
s11	0.85993	0.13311	0.19537	0.23109
s16	-0.05189	-0.01933	0.93867	0.01124
s17	0.18318	0.08013	0.39230	0.06198
s19	0.95533	0.11795	0.02426	0.00355
s20	0.02494	0.22701	0.12750	0.69453
s21	-0.00524	0.94072	0.14756	0.02454
s22	0.12785	0.61550	0.05700	0.70836
s23	0.54507	0.68496	-0.06010	0.22862
s24	0.35358	0.78151	0.21250	0.30789
s29	0.97289	0.11838	0.07406	0.07706
s31	0.59094	0.14631	0.40677	0.16321

Variance explained by each factor			
Factor1	**Factor2**	**Factor3**	**Factor4**
4.2807159	2.5037716	2.2204329	2.0584215

Item communality (h^2): Total = 11.063342							
s3	**s4**	**s10**	**s11**	**s16**	**s17**	**s19**	**s20**
0.90352402	0.88277088	0.82822383	0.84877181	0.88428782	0.19771488	0.92716240	0.55077815

s21	**s22**	**s23**	**s24**	**s29**	**s31**
0.90735313	0.90021642	0.82213962	0.87572683	0.97195881	0.56271330

I highlighted the meaningful loadings (greater than 0.60) and red marked the loadings that were meaningful for more than one items and the items without a meaningful factor loading. I found that all the items were explained by some factors, with the exception of the groups S31 and S17 whose variance was only partially explained by factor 1 (factor loading, 0.59) and factor 3 (factor loading, 0.39), respectively. The group S22 had meaningful loadings for both factor 2 and factor 4.

Apart from these weaknesses that could be overcome (for example, S31 could be considered as explained by factor 1), it was difficult to interpret the meaningfulness of the four factors. In fact, it easy to see also from Figures 4 (Appendix 3) that the new grouping "proposed" by FA was not related to the class to which flavonoids belonged, neither to the foods from which flavonoids derived. The following table shows flavonoids by group and by factor with the corresponding major food sources. Factor 1 included anthocyanidins, a flavanol, a flavone and proanthocyanidins, factor 2 represented flavonols (but not all flavonols), factor 3 included anthocyanidins and a flavanone, and factor 4 included flavanols, a flavone and a flavonol. Since each factor explains the variance in flavonoids deriving from more than one food and the same food was included in different factors, it is difficult to indicate a major food as representative of one factor. However, I found that wine was an important source for factor 1, various vegetables for factor 2, red fruit for factor 3 and tea for factor 4, as the following table shows.

Factor	Group	Flavonoids	Major sources
Factor 1	S4	FL4, ANTHO-Delphinidin	wine
		FL5, ANTHO-Malvidin	wine
		FL7, ANTHO-Peonidin	wine
		FL8, ANTHO-Petunidin	wine
		FL9 FLAVAN3-(+)-Catechin	wine
	S11	FL11, FLAVAN3-(-)-Epicatechin	apple
	S19	FL19, FLAVONE-Apigenin	various vegetables
	S29	FL29, PROANTHO-Monomers	wine
		FL30, PROANTHO-Dimers	wine
Factor2	S21	FL21, FLAVONO-Isorhamnetin	onion, mix salad
	S22	FL22, FLAVONO-Kaempferol	mix salad/various vegetables
	S23	FL23, FLAVONO-Myricetin	wine
	S24	FL24, FLAVONO-Quercetin	apple/all
Factor3	S3	FL3, ANTHO-Cyanidin	strawberry/cherry
		FL6 ANTHO-Pelargonidin	strawberry/cherry
	S16	FL 16, FLAVA-Eriodictyol	fruit juice/jam
Factor4	S10	FL 10, FLAVAN3-(-)-Epigallocatechin	tea
		FL12, FLAVAN3-(-)-Epicatechin 3-gallate	tea
		FL13, FLAVAN3-(-)-Epigallocatechin 3-gallate	tea
		FL14, FLAVAN3-Theaflavin	tea
		FL 15, FLAVAN3-Thearubigins	tea
		FL25, FLAVAN3-Theaflavin-3,3'-digallate	tea
		FL26, FLAVAN3-Theaflavin-3'-gallate	tea
		FL27, FLAVAN3-Theaflavin-3-gallate	tea
		FL28, FLAVAN3-(+)-Gallocatechin	tea
	S20	FL20, FLAVONE-Luteolin	spinach/greens
	S22	FL22, FLAVONO-Kaempferol	salad/tea/other vegetables

The following tables show the results of the solution with six factors after Varimax rotation.

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
s3	0.05035	0.16712	0.00745	0.96312	0.10531	0.07554
s4	0.97793	0.07832	-0.07646	-0.01092	0.02718	0.00976
s10	0.08920	-0.03238	0.93169	-0.01717	-0.06295	-0.00294
s11	0.69572	0.11637	0.18285	0.05039	0.61827	0.07633
s16	-0.01754	-0.00418	0.03592	0.97896	0.05654	0.07200
s17	0.08757	0.07270	0.02819	0.13002	0.10866	0.97829
s19	0.97224	0.09319	0.03889	0.03263	0.13485	0.05973
s20	-0.07014	0.22549	0.65974	0.04819	0.29002	0.02296
s21	0.00465	0.94277	0.02678	0.12062	0.05535	0.02740
s22	0.14650	0.60691	0.72803	0.05549	0.01790	0.05430
s23	0.61242	0.66637	0.27552	-0.02846	-0.04874	0.04846
s24	0.23762	0.77605	0.26619	0.08177	0.44091	0.08168
s29	0.91466	0.09547	0.07888	0.02115	0.34520	0.05358
s31	0.31737	0.14276	0.05688	0.16075	0.88149	0.11452

Variance explained by each factor					
Factor1	**Factor2**	**Factor3**	**Factor4**	**Factor5**	**Factor6**
3.7993355	2.4464412	2.0332080	1.9606274	1.6116935	1.0066740

Item communality (h^2): Total = 12.857980							
s3	**s4**	**s10**	**s11**	**s16**	**s17**	**s19**	**s20**
0.97491493	0.96927215	0.88132343	0.92162955	0.96836367	0.99951312	0.97826833	0.57798445

s21	**s22**	**s23**	**s24**	**s29**	**s31**
0.90791854	0.92616910	0.90054915	0.93734282	0.97441558	0.94031477

The total variance explained by six factors was equal to 91.8% (=12.86/14*100) and all flavonoids were related to at least one factor, but some factors were strongly correlated with the same group of flavonoids (factor1 and factor 2 with S23 and factor 2 and factor 3 with s22) making again the solution difficult to interpret. Moreover, factor 6 explained only the variance of group S17 and this was not performing. I then decided to analyze also the solution with five factors: the results follow.

	Factor1	Factor2	Factor3	Factor4	Factor5
s3	0.06204	0.16488	0.00941	0.95951	0.15072
s4	0.97226	0.09927	-0.04150	0.01004	0.00049
s10	0.09554	-0.02572	0.92085	-0.00997	-0.09806
s11	0.78637	0.12178	0.19661	0.03374	0.45775
s16	-0.00942	-0.00628	0.03634	0.97683	0.11231
s17	0.03826	0.06511	0.00840	0.13003	0.68871
s19	0.97369	0.10590	0.02076	0.03140	0.05752
s20	-0.02576	0.21975	0.68277	0.03649	0.24761
s21	-0.00113	0.94134	0.02736	0.11951	0.07643
s22	0.14661	0.61183	0.71919	0.05915	0.01713
s23	0.58534	0.67750	0.25236	-0.02310	-0.06419
s24	0.29702	0.77578	0.27989	0.06994	0.37352
s29	0.95191	0.10601	0.07435	0.01286	0.22029
s31	0.45407	0.13866	0.09498	0.13350	0.72739

Variance explained by each factor				
Factor1	**Factor2**	**Factor3**	**Factor4**	**Factor5**
4.0915904	2.4680907	2.0309843	1.9366099	1.5208082

Item communality (h^2): Total = 12.048084							
s3	**s4**	**s10**	**s11**	**s16**	**s17**	**s19**	**s20**
0.97449981	0.95696837	0.86747337	0.88252923	0.96825190	0.49700754	0.96401566	0.57776588

s21	**s22**	**s23**	**s24**	**s29**	**s31**
0.90700376	0.91686022	0.86997540	0.91279765	0.97159973	0.78133508

The total variance explained by five factors was equal to 86% (=12.04/14*100), all flavonoids were related to at least one factor, only one item was correlated with two factors (S22 with factor1 and factor 2), and there was no factor correlated only with one item. This solution appeared then more convincing. Factor 1 and factor 2 were comparable to the same factors of the solution with four factors whereas factors 3 and factor 4 were interchangeable with the previous solution. In addition,

factor 5 explained the variance of those flavonoids that were not correlated to any factors in the solution with four factors., that is flavanones - mainly deriving from citrus fruit - and proanthocyanidins with three or more mers - mainly deriving from apples (Figure 3 - Appendix 3). This made the solution of five factor the best one for our aim to describe the entire "construct flavonoids".

This solution allows to study possible effects of interaction between flavonoids on cancer risk, besides evaluating the individual effect of a flavonoid or a class of flavonoids as previous studies did. I therefore examined the relation between the five factors and colorectal cancer risk.

Figure 5 (Appendix 3) shows the ORs for colorectal cancer risk of the second and the third tertiles versus the lowest tertile of factors. I found that factor 2, factor 4 and factor 5 were inversely associated to the risk of colorectal cancer, whereas no meaningful association was found for factor 1 and factor 3. The OR for highest versus the lowest tertile was 0.66 for factor 2, 0.84 for factor 4, and 0.85 for factor 5. The trend in risk was significant in all three factors also when considering quintiles instead of tertiles.

Factor	Group	Flavonoids	Major sources
Factor2	S21	FL21, FLAVONO-Isorhamnetin	onion, mix salad
	S22	FL22, FLAVONO-Kaempferol	mix salad/various vegetables
	S23	FL23, FLAVONO-Myricetin	wine
	S24	FL24, FLAVONO-Quercetin	apple/all
Factor4	S3	FL3, ANTHO-Cyanidin FL6 ANTHO-Pelargonidin	strawberry/cherry strawberry/cherry
	S16	FL 16, FLAVA-Eriodictyol	fruit juice/jam
Factor5	S17	FL17, FLAVA-Hesperetin FL18, FLAVA-naringenin	citrus fruit citrus fruit
	S31	FL31, PROANTHO-Trimers FL32, PROANTHO-4-6mers FL33, PROANTHO-7-10mers FL34, PROANTHO-Polymers	apple apple apple apple

6.1.2 Factor analysis on 8 classes of flavonoids and proanthocyanidins

I report here the results of FA applied to the 8 classes of flavonoids instead of the single flavonoids. As mentioned before, this method does not take into consideration the different variances in flavonoids belonging to the same classes, that we observed before.

The total MSA was 0.57 with a minimum individual MSA of 0.34 (for isoflavones). I chose the solution with 4 factors according to the eigenvalue >1 and the cumulative percentages of variance extracted (82%) greater than 75% criteria.

Eigenvalues of correlation matrix: Total = 8 Mean = 1				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	3.29213112	2.08540145	0.4115	0.4115
2	1.20672967	0.18122097	0.1508	0.5624
3	1.02550870	0.02785415	0.1282	0.6905
4	0.99765455	0.39757231	0.1247	0.8153
5	0.60008223	0.13252372	0.0750	0.8903
6	0.46755852	0.11099813	0.0584	0.9487
7	0.35656039	0.30278556	0.0446	0.9933
8	0.05377483		0.0067	1.0000

After Varimax rotation, I obtained the following factors loading:

	Factor1	**Factor2**	**Factor3**	**Factor4**
iso	-0.00349	-0.00208	-0.00391	0.99877
antho	0.92454	0.07478	-0.05543	-0.01782
flavan3	0.39486	0.72357	-0.08913	-0.02528
flava	-0.00462	0.03142	0.94873	-0.01094
flavone	-0.04212	0.91551	0.05338	-0.00220
flavono	0.39272	0.65785	0.21065	0.02855
prop	0.94942	0.23743	0.01525	-0.00631
prott	0.59144	0.34748	0.40661	0.05041

Variance explained by each factor			
Factor1	**Factor2**	**Factor3**	**Factor4**
2.4179174	1.9781806	1.1239148	1.0020112

Item communality (h^2): Total = 6.522024							
iso	**antho**	**flavan3**	**flava**	**flavone**	**flavono**	**prop**	**prott**
0.99756576	0.86376143	0.68804949	0.90122503	0.84279159	0.63218465	0.95803527	0.63841082

As expected, the variances in isoflavones and flavanones were explained each by one single factor not related to any other class. In fact, the maximum correlation coefficient with other classes of flavonoids was < 0.1 for isoflavones and 0.2 for flavanones. This could be not a performing solution, but when I applied the FA to all the classes of flavonoids excluding isoflavones and flavanones, I found that the best solution was with 2 factors, which were identical to the factor 1 and factor 2 of this analysis. For this reason, I remained with this solution and estimated the ORs according to tertiles of factors.

I found that only factor 4 was inversely associated to colorectal cancer risk: the ORs of the second and third tertiles were 0.83 and 0.74, respectively. This solution appeared poor compared to the solution that I found by inserting single flavonoids.

6.2 Collinearity between flavonoids and wine in the oral and pharyngeal cancer data

6.2.1 Computing the intake of flavonoids not deriving from wine

I am interested to detect which flavonoids are related to the disease taking into account the possible influence of their major food source on the risk. The simplest way to address the problem of the high correlation between flavonoids and the food is to compute the intake of flavonoids excluding the intake deriving from the consumption of this food. Among major flavonoid food sources, the high consumption of wine in the Italian population gave a lot of problems in analyzing flavonoids and the risk of some cancers that are strongly related to alcohol consumption.

Here, I report an example for oral and pharyngeal cancer. I used data from the Italian case-control on oral and pharyngeal cancer, and analysed monomer and dimer proanthocyanidins and proanthocyanidins with three or more mers that are strongly contained in wine: seventy-five per cent of monomers and dimers and 25% of polymers with three or more mers derived by wine in our controls (Figure 6 - Appendix 3). I used a logistic model that included also alcohol intake (*see* Material and Method), and I found that the ORs for proanthocyanidins with and without the intake from wine were different (Figure 7 - Appendix 3). In particular, I found significant associations between proanthocyanidins without intake from wine and oral and pharyngeal cancer risk (OR, 0,64 for monomers and dimers combined and OR, 0,71 for polymers with 3 or more mers), that were not significant when I considered also intake from wine.

6.2.2 A FA application

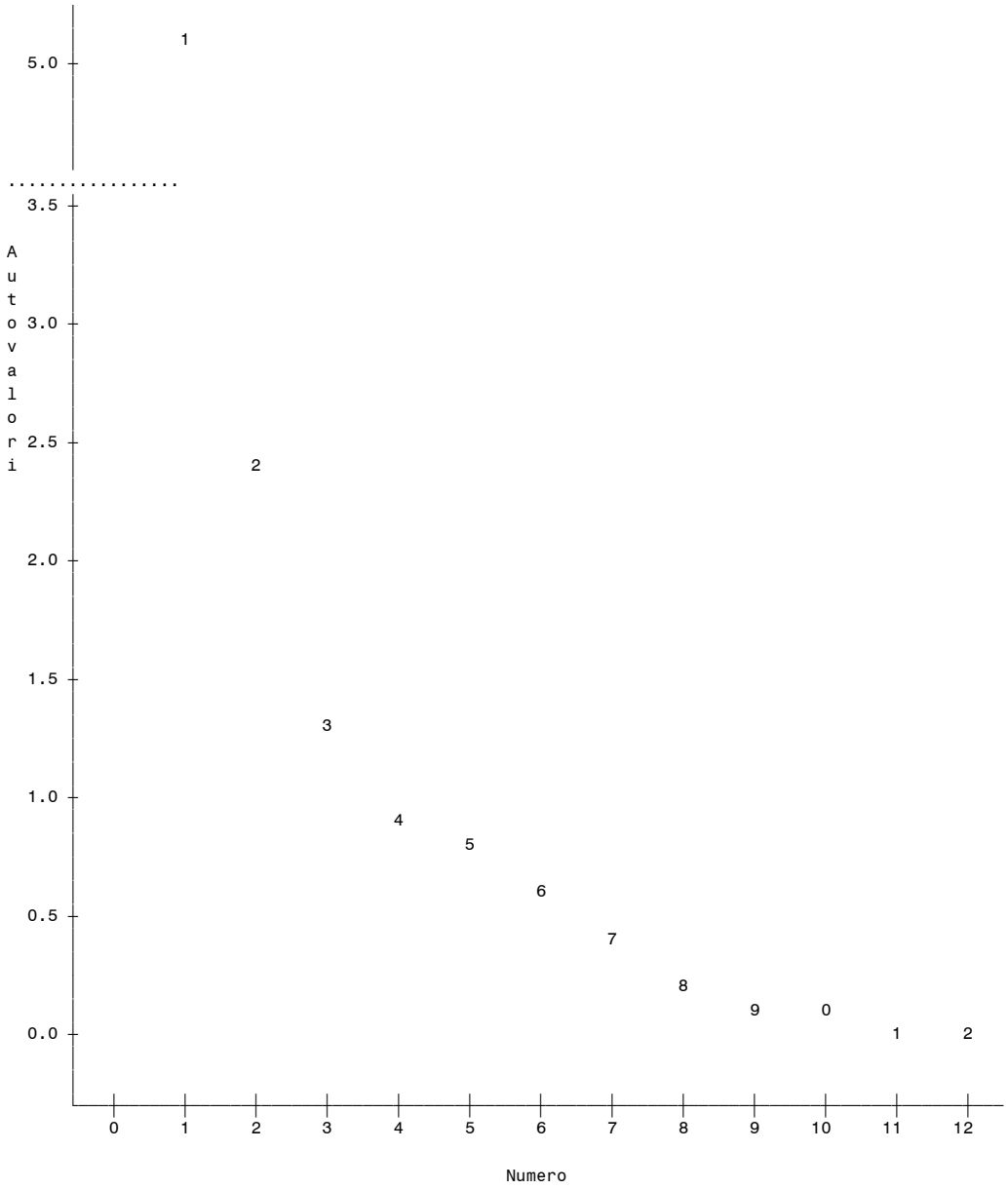
I applied FA by excluding from the beginning isoflavones and flavanones since their maximum correlation coefficients with other flavonoids was 0.2 (for FLAVA-Hesperetin with FLAVONO-Quercetin). I inserted one flavonoid as representative per one group according to the first classification I did in the FA on 34 flavonoids (Figure 3 - Appendix 3), for a total of 12 flavonoids. The total MSA was quite low, 50%.

I chose the solution with 3 factors according to the eigenvalue >1 and the scree plot observation criteria.

Eigenvalues of correlation matrix: Total = 12 Mean = 1				
	Autovalore	**Differenza**	**Proporzione**	**Cumulativa**
1	5.05121425	2.65595497	0.4209	0.4209
2	2.39525929	1.05400988	0.1996	0.6205
3	1.34124941	0.39335237	0.1118	0.7323
4	0.94789704	0.13948759	0.0790	0.8113
5	0.80840945	0.19381226	0.0674	0.8787
6	0.61459719	0.17246928	0.0512	0.9299
7	0.44212791	0.26924642	0.0368	0.9667
8	0.17288149	0.03893569	0.0144	0.9811

Eigenvalues of correlation matrix: Total = 12 Mean = 1				
	Autovalore	**Differenza**	**Proporzione**	**Cumulativa**
9	0.13394580	0.06999604	0.0112	0.9923
10	0.06394976	0.03597513	0.0053	0.9976
11	0.02797463	0.02748085	0.0023	1.0000
12	0.00049378		0.0000	1.0000

Scree plot



The cumulative percentages of variance extracted was 73%. After Varimax rotation, I obtained the following factors loading:

		Factor1	Factor2	Factor3
FL3	ANTHO-Cyanidin	0.08597	-0.05047	0.53572
FL4	ANTHO-Delphinidin	0.93233	-0.05900	-0.03097
FL10	FLAVAN3-(-)-Epigallocatechin	0.06783	0.89961	-0.17177
FL11	FLAVAN3-(-)-Epicatechin	0.92925	0.13477	0.13998
FL19	FLAVONE-Apigenin	0.14191	0.78343	0.00946
FL20	FLAVONE-Luteolin	-0.03924	0.65827	0.22353
FL21	FLAVONO-Isorhamnetin	0.06836	0.19502	0.86956
FL22	FLAVONO-Kaempferol	0.18645	0.83326	0.36436
FL23	FLAVONO-Myricetin	0.77157	0.29180	0.33391
FL24	FLAVONO-Quercetin	0.48820	0.37841	0.69061
FL29	Proanthocyanidin-Monomers	0.97599	0.09539	0.01032
FL31	Proanthocyanidin-Trimers	0.63749	0.03608	0.27023

Variance explained by each factor		
Factor1	**Factor2**	**Factor3**
3.9984566	2.8516824	1.9375840

Item communality (h^2): Total = 8.787723							
FL3	**FL4**	**FL10**	**FL11**	**FL19**	**FL20**	**FL21**	**FL22**
0.29693548	0.87368405	0.84340436	0.90126614	0.63399627	0.48482703	0.79884468	0.86184960

FL23	**FL24**	**FL29**	**FL31**
0.79196192	0.85847874	0.96175717	0.48071752

It is interesting to see that all and only the flavonoids contained in wine load significantly to the first factor, which explained 33.3% of the total variance and seems to condense the variance of wine consumption. In fact, the groups of flavonoids strongly correlated to factor 1 were:

- FL4, ANTHO-Delphinidin ~ FL5, ANTHO-Malvidin ~ FL7, ANTHO-Peonidin, FL8, ANTHO-Petunidin ~ FL9 FLAVAN3-(+)-Catechin
- FL11, FLAVAN3-(-)-Epicatechin

- FL23, FLAVONO-Myricetin
- FL29, PROANTHO-Monomers ~ FL30, PROANTHO-Dimers
- FL31, PROANTHO-Trimers ~ FL32, PROANTHO-4-6mers ~ FL33, PROANTHO-7-10mers ~ FL34, PROANTHO-Polymers

The other 2 factors are described below and explained respectively the 23.7% and 16.1% of the total variance.

Factor	Flavonoids	Major sources
Factor2	FL 10, FLAVAN3-(-)-Epigallocatechin	tea
	FL12, FLAVAN3-(-)-Epicatechin 3-gallate	tea
	FL13, FLAVAN3-(-)-Epigallocatechin 3-gallate	tea
	FL14, FLAVAN3-Theaflavin	tea
	FL 15, FLAVAN3-Thearubigins	tea
	FL25, FLAVAN3-Theaflavin-3,3'-digallate	tea
	FL26, FLAVAN3-Theaflavin-3'-gallate	tea
	FL27, FLAVAN3-Theaflavin-3-gallate	tea
	FL28, FLAVAN3-(+)-Gallocatechin	tea
	FL19, FLAVONE-Apigenin	various vegetables
	FL20, FLAVONE-Luteolin	spinach/greens
	FL22, FLAVONO-Kaempferol	salad/tea/other vegetables
Factor3	FL3, ANTHO-Cyanidin	strawberry/cherry
	FL6 ANTHO-Pelargonidin	strawberry/cherry
	FL21, FLAVONO-Isorhamnetin	onion, mix salad
	FL24, FLAVONO-Quercetin	apple/all

These 2 factors were slightly correlated to other flavonoids (e.g., the factors loading were around 0.3 between factor 3 and FL22, FLAVONO-Kaempferol, FL23, FLAVONO-Myricetin and proanthocyanidins $\geq$ 3 mers), but were totally purified by the confounding of wine consumption.

I computed the ORs for the oral and pharyngeal cancer risk according to tertiles of factors 1-3, using the previous model. Only the factor 3 was inversely related to the risk of this tumor, as shown in Figure 8 (Appendix 3). The OR for the highest versus the lowest tertile was 0.66 (95% CI, 0.51-0.87). I computed also the ORs according to quintile of factors and obtained comparable results.

Figure 9 – a) (Appendix 3) shows the distribution of subjects by levels of factor 1 and factor 3. There was not a clear trend although the χ^2 test was significant. I obtained a similar result when I studied the distribution of subjects by levels of alcohol intake and factor 3 (Figure 9 - b). Figures 10

a) and b) (Appendix 3) show the ORs for oral cavity and pharyngeal cancer according to the various combination of levels of factor 1 and factor 3, and of alcohol habits and factor 3, comparing to the lowest levels. Test for interaction was not significant. However, this result shows better the effect of a diet rich in flavonoids in heavy alcohol drinkers, since the OR was 13 in high flavonoid consumer and 7 in low flavonoids consumers. These analyses are difficult using classical methods since high levels of alcohol intake were related to high levels of flavonoid intake.

I obtained similar results when I added to these flavonoids an item equal to the sum of all flavonoids deriving from wine consumption, or directly an item for alcohol intake. I applied also the FA to all flavonoids without intake deriving from wine, but the results were difficult to interpret and to use.

6.3 Collinearity between flavonoids and other biocative compounds contained in plant foods

6.3.1 Residuals on TAC

It is not easy to take into account the possible influence of other beneficial compounds contained in the same food sources of flavonoids. Among them, however, all the components with antioxidant properties can be measured trough the TAC, using the three assays TEAC, TRAP and FRAP. Also flavonoids contribute to the computation of these measures, together with other antioxidants like vitamin A and vitamin C. I computed then the residuals of flavonoids on TAC in order to evaluate the risk for high levels of flavonoids independently from high levels of TAC. I applied this technique to oral and pharyngeal cancer data.

Given the high correlation between the three measures of TAC, I will show the results only for FRAP, which was inversely associated to the risk of oral and pharyngeal cancer. In Figure 11 (Appendix 3), I reported the ORs for the highest versus the lowest quintile of flavonoid intake and residuals of flavonoids on FRAP for those classes that were inversely related to this tumor. I found that the association for flavanones and flavonols persisted after adjusting for FRAP, whereas the

association for proanthocyanidins with three or more mers disappeared. This was an interesting result, especially for flavanones, since this class was strongly correlated to vitamin C ($r \sim 0.8$), which was inversely related to oral and pharyngeal cancer risk too. The models including both flavanones and vitamin C were difficult to interpret, and the residual method allowed me to evaluate the risk of flavanones adjusting – though indirectly - for vitamin C.

This result strengthens the hypothesis that flavanones and flavonols – but not proanthocyanidins - have a role in the oral and pharyngeal cancer prevention.

7. RESULTS - A METHOD TO TAKE INTO ACCOUNT THE OVER-REPORTING OF CASES IN CASE-CONTROL STUDIES

7.1 Residuals on fruit and vegetable consumption in the oral and pharyngeal cancer data

A protective role of fruit and vegetable consumption on various cancer risk has been reported in numerous case-control studies but results on cohort studies were open to discussion [1]. The results from case-control studies can be in part explained by a recall bias of cases, who tend answer with more accuracy and report a higher intake of energy. Likely, the weak associations that I found in these studies reflect this bias, whereas the stronger associations, which I found especially for proanthocyanidins, show a more convincing protective effect of these compounds on cancer. To investigate this issue, I computed the residuals of flavonoids on fruit and vegetable consumption in all subjects of oral and pharyngeal cancer study, as a direct adjustment of a recall bias of cases. Figure 1 (Annex 4) shows the ORs for oral cavity and pharyngeal cancer, according to the highest versus the lowest quintile of flavonoid residuals on fruit and vegetables. No associations persisted after this adjustment.

7.2 Residuals on water consumption in the oral and pharyngeal cancer data

Although water does not differ in terms of cancer risk from case-control and cohort studies, like fruit and vegetables, the use of residuals on water rather than energy intake was considered to take into account the bias due to the over-reporting of cases in case-control studies.

I calculated the intake of water from food and beverages using food database tables and information from our FFQ, and computed residuals on water in all subjects of oral and pharyngeal study. The ORs are reported in Figure 1 (Annex 4). The association between flavanones and flavonols did not substantially change, whereas the association for proanthocyanidins with three or more mers disappeared.

8. DISCUSSION

8.1 Flavonoids and cancer risk

Flavanones, which are flavonoids deriving mainly from citrus fruit, were the class most strongly associated with a reduced risk of neoplasms at the upper aerodigestive tract in these Italian studies. They have also been inversely related to gastric cancer in a Greek case-control study [14]. This is of interest, given the similarities in risk factors between various neoplasms of the upper aerodigestive tract (mainly tobacco, alcohol, as well as a diet poor in vegetables and fruit) [50, 51], and gastric cancer, too (i.e., tobacco, lower social class and various indicators of a poor diet) [52]. As was noted for esophageal and laryngeal cancers, the inverse relation between flavanones and stomach cancer persisted, albeit weaker, after allowance for vitamin C [14, 40, 41]. Data from Italy are also in agreement with those of a case control-study conducted in Uruguay that reported inverse relations between flavonoids and oral, esophageal, and laryngeal cancer risk. No specific information on types of flavonoids was given in that study. More recently, in a population-based case-control study from the USA [53], inverse associations in white men were observed between anthocyanidin intake and esophageal adenocarcinoma, and isoflavone intake and esophageal squamous cell carcinoma. However, none of these associations remained significant after adjusting for dietary fiber. To our knowledge, no other study investigated flavonoids in relation to the risk of cancer of the upper aerodigestive tract.

Inverse associations between proanthocyanidins and cancer risk were found for pancreatic cancer. Further adjustment for fruit and vegetables, or vitamin C, did not materially change these associations. proanthocyanidins with three or more mers, which derived mainly from apples, were most strongly related to the risk of this cancer. These associations did not substantially change after adjustment for total flavonoid intake. There was little evidence that other flavonoids had a significant role on stomach cancer risk.

These results are in line with findings from previous studies *in vitro* [54] that showed superior radical scavenging properties of proanthocyanidins as compared with other flavonoids. proanthocyanidins may lead to oligomerization via phenolic coupling and enlargement of the number of reactive sites.

This is the first epidemiological study to suggest that dietary proanthocyanidins have a favourable role on gastric cancer risk.

With reference to colorectal cancer, a case-control study from Canada indicated an inverse association between dietary isoflavone intake and risk of colorectal cancer [55]. In a prospective study from Japan [56], intake of isoflavones (and soy foods) was inversely associated with the risk of proximal colon cancer in men, while another Japanese prospective study [57] showed a reduced risk of colon cancer with increased consumption of soy products (major sources of isoflavones) only among women. No other consistent associations, however, were reported from these Japanese prospective studies undertaken in populations characterized by a high consumption of isoflavone-rich foods [56, 57]. In a case-control study in Japan, isoflavone intake has also been inversely associated to the risk of colorectal adenoma [58], and, together with flavonol intake, to the risk of advanced adenoma recurrence. A recent case-control study conducted in Sweden indicated a significant decreased risk of colorectal cancer for intake of anthocyanidins and flavonols [59], but not for isoflavones and flavones, as in the Italian data. In the Italian population, anthocyanidins were derived mainly from wine, red fruit, and onions, and flavonols from apples or pears, wine and mixed salads.

I found that also proanthocyanidin intake was inversely associated with the risk of colorectal cancer. The inverse associations were apparently stronger for the classes of proanthocyanidins with higher degree of polymerization. A recent case-control study from Scotland found an inverse association of colorectal cancer risk with the intake of single flavanols, including catechins and epicatechins, and procyanidins (polymers of [epi] catechin) [60], but no other study has

systematically investigated proanthocyanidins. Studies in vitro and experimental animals suggest they have favorable effects on colorectal cancer [4, 61-67] and have larger antioxidant effects than flavanols [68, 69]. In this analysis, adjustment for flavanols and anthocyanidins, which were strongly correlated to proanthocyanidins, did not substantially modify the inverse relation between proanthocyanidins and colorectal cancer risk.

Intake of flavones – but not of other flavonoids – was inversely related to HCC risk. Moreover, total flavonoids, flavan-3-ols and anthocyanidins may be strongly inversely associated with CAC. No epidemiological study has previously investigated the relation between flavonoids and primary liver cancer risk, and larger studies are needed for documentation of the strength of these associations.

Isoflavones and proanthocyanidins with three or more mers were inversely related to the risk of pancreatic cancer, also after adjustment for fruit and vegetables consumption, and for vitamin C and folate intakes. Less strong associations were also found for flavanone and flavonol intakes that, however, disappeared after adjustment for other dietary micronutrients. The OR estimates were generally lower for women than men. This may be due to the fact that men drink more wine (a main source of flavonoids and PAs) than women, and this can reduce the inverse association between flavonoids and proanthocyanidins and pancreatic cancer in men, given the direct association between alcohol consumption and pancreatic cancer risk.

In the four cohort studies that examined the relation between flavonoids and pancreatic cancer risk [70-73], only flavanols, flavones and flavonols have been studied. An association between flavonols (in particular, kaempferol) and pancreatic cancer was reported only in subgroups of population [70, 71]. The risks, however, were not adjusted for other dietary factors.

With reference to breast cancer, these findings are compatible with those of a case-control study from Greece [17], which found an inverse association between flavones and breast cancer risk. A recent case-control study from the USA on 1434 women of Long Island [16] reported results in line

with those from the Mediterranean populations. Among Italian women, flavone intake derived mainly from (aromatic) herbs, and flavonol intake from various common vegetables and fruits. Several epidemiological studies also found that dietary phytoestrogens are inversely associated with breast cancer risk [74]. The absence of any meaningful association with isoflavone intake in the Italian study may be due to the extremely limited intake of soya or soya products - and consequently of isoflavones - in the Italian population.

These results on ovarian cancer are also in line with the litterature. The intake of isoflavones was associated with lower risk of ovarian cancer in a US cohort study [19], as well as in a Chinese case-control study [75]. An inverse association between isoflavones and ovarian cancer risk can be explained by the observation that isoflavones have anti-estrogenic effects [5], and hence may inhibit the growth and proliferation of ovarian cell lines [76-78]. Another cohort study from the US reported a significant decrease in ovarian cancer incidence for the highest versus the lowest quintile of flavonol kaempferol [79], in line with these results.

Although some flavonoids have been reported to have a favourable effect against prostate cancer [80], the results from epidemiological studies are inconsistent. A recent prospective study from Japan found that isoflavone intake was associated with a decreased risk of localized prostate cancer [18]. If confirmed, this finding could partly explain the lower incidence of prostate cancer in Asian as compared to Western populations characterized by a low consumption of isoflavone-rich foods as soya. In the Italian study, isoflavones derive mainly from beans, soy and soy products, the consumption of which is limited in Italy, and this may explain the inconsistent relation observed with prostate cancer. In a case-control study nested in the European Prospective Investigation into Cancer and Nutrition EPIC cohort study [81], higher plasma concentrations of isoflavone genistein, but not other isoflavones, were associated with lower risk of prostate cancer.

With reference to kidney cancer, a recent study showed an inverse association between the flavonol quercetin and renal cancer among smokers of the Alpha-Tocopherol Beta-Carotene Cancer Prevention (ATBC) Study cohort [82], in line with these results.

8.2 TAC and cancer risk

I obtained the estimates of dietary TEAC, TRAP and FRAP that were comparable with previous values showed in the literature [83]. Fruit and vegetable mainly contributes to the TEAC (42%), TRAP (35%) and FRAP (40%) intakes. Among the principal food sources, there were wine, citrus fruits, apples and pears, and bread. Given the high correlation between the three TAC assays, examining only one measure should be considered for the next studies.

For colorectal cancer, we found that TEAC, TRAP and FRAP were inversely related with the risk of this tumor. Associations were stronger for rectal cancer. These findings are in agreement with those from the Health Professional Follow-up Study that found an inverse association between FRAP and rectal cancer (OR=0.58, 95% CI, 0.35-0.96) on a cohort of 47,399 men including 201 rectal cancers [20].

Adjustment for flavonoids and proanthocyanidins reduced the strength of the inverse relations between TAC and colorectal cancer risk. The association disappeared after adjusting for anthocyanidin intake. Given the high correlation between TAC and anthocyanidins, however, these results are difficult to be interpreted.

Availability of TAC estimates allowed me to investigate the associations that I found between anthocyanidins, flavones, flavonols, and proanthocyanidins with more than 10 mers and colorectal cancer, further adjusted for TAC intakes. I found that ORs did not change after allowance for FRAP. I obtained similar OR estimates adjusting for TEAC and TRAP rather than FRAP.

8.3 Statistical techniques to face the problem of high collinearity among dietary factors

FA allows to study whether there are possible patterns of flavonoids more associated to the risk of cancer than single flavonoids or classes of flavonoids. We found that the variance of the entire structure of flavonoids can be explained by five factors, and found that 3 of them were inversely related to colorectal cancer risk. Based on this result, it is possible to define a pattern of foods in order to have a more practical results, also in terms of dietary modification recommendations for cancer prevention using a multivariate regression model or a reduced rank regression. This would allow to find a cluster of foods that maximizes the variance of these factors and has a protective effect on the disease trough flavonoids.

In nutritional epidemiological research, dietary patterns have substantially used over the past several years [84]. Patterning methods consider multiple foods, beverages, and/or nutrients and create dietary variables that more realistically resemble actual eating behavior, but no study have investigated specifically flavonoids, neither included flavonoids in the analyses. Only one study investigated a dietary pattern that was predictive of flavonol intake and pancreatic cancer risk in a German study [85]. Using data from the Multiethnic cohort, Nothlings and colleagues found a flavonol based food pattern, consisting mainly of tea, fruit, cabbage, and wine, that was inversely associated to the risk of pancreatic cancer in that population; they could not, however, replicate this association in the EPIC study [85]. I also identified a combination of foods from the flavonoids strongly related to pancreatic cancer risk. It was different from that suggested in the German study, but it was inversely associated to pancreatic cancer. Eating one more portion of apples, pears, pulses, peaches, apricots or prunes every day reduced the risk of pancreatic cancer by 25% [45].

The definition of dietary patterns on all complex of nutrients and dietary compounds including flavonoids would allow to describe in a complete manner all the variances from the diet that play a role in the evaluation of cancer risk. This work was started by the Department of Occupational and Environmental Health in 2008 [86], but did not include flavonoids up to now.

All these analysis could be applied to residuals on energy intake. I did adjust for energy intake in the logistic models, but in some cases the use of residuals is preferable for a more strong adjustment. Residuals would have allowed me to easily compare cases and controls and selected covariates (e.g. sex, age) by levels of flavonoid intake (and their factors) taking into account differences in energy intake. However, more difficulties will come up in adjusting for wine consumption since the residuals on energy includes also the adjustment for alcohol intake which contributes to the total energy intake.

Computing the intake of flavonoids without the intake deriving by wine consumption allowed me to detect significant associations with oral and pharyngeal cancer that I would not have detected using total intake. However, this method does not allow to evaluate the effect that the total amount of flavonoids, taken by the diet, has on cancer risk. There is not a biological explanation for which flavonoids deriving by wine have not protective effect on the tumour, but is only a statistical problem in being able to eliminate the confounding effect of the consumption of wine on the risk. FA has resolved in part this problem and has allowed to make further analysis in strata.

There are flavonoid for which their intake derives exclusively by one and only one source of food. For example, FLAVAN3-(-)-Epigallocatechin derives only by tea (99.0%), or FLAVA-Hesperetin derives only by citrus fruit (92.8%). In this case, we must consider that the flavonoid and the food source have a common variance and it is not possible disentangle their effects on the risk of disease.

8.4 A method to take into account the over-reporting of cases in case-control studies

It is difficult to interpret and compare the results on the residuals on fruit and vegetables and on the residuals on water. Of course, water can be used as energy in adjusting for a recall bias of a case-control study design, and has not the limitation to be related to the cancer risk like fruit and vegetables, or energy. In fact, the use of residuals on fruit and vegetables is a strong adjustment and can weaken the risk estimates of flavonoids.

It interesting to see, however, that the residual method on TAC gave the same results of the residuals on water. In both analysis in fact, the relations for flavanones and flavonols persisted whereas the association for proanthocyanidins disappeared. These analyses provide evidence of a protective effect of flavanones and flavonols on oral and pharyngeal cancer.

9. CONCLUSION

In conclusion, the findings this project on a large network of Italian case-control studies provide support for an apparent protective role of flavanones on upper aerodigestive tract cancers, flavonols, proanthocyanidins on stomach and pancreatic cancer, anthocyanidins and proanthocyanidins on colorectal cancer, flavonols and flavones on breast cancer, isoflavones on ovarian cancer, and flavonols on renal cancer. For most investigated neoplasms, adjustment for flavonoids reduced the strength of the inverse association between vegetables or fruit consumption and the risk of cancer, whereas allowance for fruit and vegetables consumption only moderately changed the observed associations with flavonoids. Misclassification may play a role, but it appears that a diet rich in fruit and vegetables does not alone account for the protections of flavonoids on the risk of several cancer sites, whereas the inverse relation of cancer with fruit and vegetables is not totally explained by flavonoid intake.

With reference to TAC, the analysis on colorectal cancer found that TAC was inversely associated to the risk of this tumor. However, a diet rich antioxidants does not alone account for the protections of flavonoids on the risk of colorectal cancer. Similar analysis on other cancer sites are needed.

The application of FA to the “construct of flavonoids” investigated the structure of flavonoids and analysed their variances. This allowed to highlight the importance to analyse individual flavonoids or groups of flavonoids instead of their classes, and to resolve – in part – the problem of the strong confounding effect of wine consumption. Moreover, these results suggest to include flavonoids and TAC in the estimation of dietary patterns in nutritional epidemiologic research.

At last, the proposal to use residuals on water or fruit and vegetables leaves open an epidemiological question, given the criticism that the case-control studies still face in the epidemiological literature, but can be of any help since new statistical methods are needed to resolve these limitations.

REFERENCES

1. World Cancer Research Fund and American Institute for Cancer Research. Food, Nutrition, Physical Activity and the Prevention of Cancer: a Global Perspective. Washington, DC: American Institute for Cancer Research 2007.
2. La Vecchia C, Bosetti C. Diet and cancer risk in Mediterranean countries: open issues. Public Health Nutr 2006; 9: 1077-1082.
3. La Vecchia C, Tavani A. Fruit and vegetables, and human cancer. Eur J Cancer Prev 1998; 7: 3-8.
4. Manach C, Scalbert A, Morand C et al. Polyphenols: food sources and bioavailability. Am J Clin Nutr 2004; 79: 727-747.
5. Ross JA, Kasum CM. Dietary flavonoids: bioavailability, metabolic effects, and safety. Annu Rev Nutr 2002; 22: 19-34.
6. La Vecchia C, Decarli A, Pagano R. Vegetable consumption and risk of chronic disease. Epidemiology 1998; 9: 208-210.
7. Trichopoulou A, Naska A, Antoniou A et al. Vegetable and fruit: the evidence in their favour and the public health perspective. Int J Vitam Nutr Res 2003; 73: 63-69.
8. Arts IC, Hollman PC. Polyphenols and disease risk in epidemiologic studies. Am J Clin Nutr 2005; 81: 317S-325S.
9. Neuhouser ML. Dietary flavonoids and cancer risk: evidence from human population studies. Nutr Cancer 2004; 50: 1-7.
10. U.S. Department of Agriculture. Iowa State University Database on the Isoflavone Content of Foods, Release 1.3, 2002. Beltsville, MD: USDA 2002.
11. U.S. Department of Agriculture. USDA Database for the Flavonoid Content of Selected Foods. Beltsville, MD: USDA 2003.
12. U.S. Department of Agriculture. USDA Database for the Proanthocyanidin Content of Selected Foods. Beltsville, MD: USDA 2004.
13. De Stefani E, Ronco A, Mendilaharsu M, Deneo-Pellegrini H. Diet and risk of cancer of the upper aerodigestive tract--II. Nutrients. Oral Oncol 1999; 35: 22-26.
14. Lagiou P, Samoli E, Lagiou A et al. Flavonoids, vitamin C and adenocarcinoma of the stomach. Cancer Causes Control 2004; 15: 67-72.
15. Lagiou P, Rossi M, Lagiou A et al. Flavonoid intake and liver cancer: a case-control study in Greece. Cancer Causes Control 2008; 19: 813-818.
16. Fink BN, Steck SE, Wolff MS et al. Dietary flavonoid intake and breast cancer risk among women on Long Island. Am J Epidemiol 2007; 165: 514-523.
17. Peterson J, Lagiou P, Samoli E et al. Flavonoid intake and breast cancer risk: a case--control study in Greece. Br J Cancer 2003; 89: 1255-1259.
18. Kurahashi N, Iwasaki M, Sasazuki S et al. Soy isoflavone consumption is not associated with increased risk of advanced prostate cancer. Cancer Epidemiol Biomarkers Prev 2007; 16: 2169.
19. Chang ET, Lee VS, Canchola AJ et al. Diet and risk of ovarian cancer in the California Teachers Study cohort. Am J Epidemiol 2007; 165: 802-813.
20. Mekary RA, Wu K, Giovannucci E et al. Total antioxidant capacity intake and colorectal cancer risk in the Health Professionals Follow-up Study. Cancer Causes Control 2010; 21: 1315-1321.
21. Gallus S, Talamini R, Giacosa A et al. Does an apple a day keep the oncologist away? Ann Oncol 2005; 16: 1841-1844.
22. Kuper H, Tzonou A, Lagiou P et al. Diet and hepatocellular carcinoma: a case-control study in Greece. Nutr Cancer 2000; 38: 6-12.

23. Franceschi S, Negri E, Salvini S et al. Reproducibility of an Italian food frequency questionnaire for cancer studies: results for specific food items. Eur J Cancer 1993; 29A: 2298-2305.
24. Decarli A, Franceschi S, Ferraroni M et al. Validation of a food-frequency questionnaire to assess dietary intakes in cancer studies in Italy. Results for specific nutrients. Ann Epidemiol 1996; 6: 110-118.
25. Gnardellis C, Trichopoulou A, Katsouyanni K et al. Reproducibility and validity of an extensive semiquantitative food frequency questionnaire among Greek school teachers. Epidemiology 1995; 6: 74-77.
26. Trichopoulou A, Katsouyanni K, Stuver S et al. Consumption of olive oil and specific food groups in relation to breast cancer risk in Greece. J Natl Cancer Inst 1995; 87: 110-116.
27. Liggins J, Bluck LJ, Runswick S et al. Daidzein and genistein content of fruits and nuts. J Nutr Biochem 2000; 11: 326-331.
28. Liggins J, Bluck LJ, Runswick S et al. Daidzein and genistein contents of vegetables. Br J Nutr 2000; 84: 717-725.
29. Liggins J, Mulligan A, Runswick S, Bingham SA. Daidzein and genistein content of cereals. Eur J Clin Nutr 2002; 56: 961-966.
30. Salvini S, Parpinel M, Gnagnarella P et al. Banca di composizione degli alimenti per studi epidemiologici in Italia. Milano, Italia: Istituto Europeo di Oncologia 1998.
31. Gnagnarella P, Parpinel M, Salvini S et al. The update of the Italian food composition database. J Food Comp Analysis 2004; 17: 509-522.
32. Trichopoulou A, Georga K. Composition tables of foods and Greek dishes. Athens, GR: Parisianou Publications 2004.
33. Pellegrini N, Serafini M, Colombi B et al. Total antioxidant capacity of plant foods, beverages and oils consumed in Italy assessed by three different in vitro assays. J Nutr 2003; 133: 2812-2819.
34. Breslow NE, Day NE. Statistical methods in cancer research. Vol. I. The analysis of case-control studies. Lyon, France: IARC 1980.
35. Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. Am J Epidemiol 1986; 124: 17-27.
36. Johnson R, Wichern D. Applied multivariate statistical analysis. In Edition 5th ed. Upper Saddler River, NJ: Prentice Hall. 2002.
37. Cattell R. The scree test for the number of factors. Multivariate Behavioral Research 1: 245-276.
38. Pett M, Lackey N, Sullivan J. Making sense of factor analysis: the use of factor analysis for instrument development in health care research. . Thousand Oaks, CA: Sage Publications; 2003. 2003.
39. Rossi M, Garavello W, Talamini R et al. Flavonoids and the risk of oral and pharyngeal cancer: a case-control study from Italy. Cancer Epidemiol Biomarkers Prev 2007; 16: 1621-1625.
40. Rossi M, Garavello W, Talamini R et al. Flavonoids and risk of squamous cell esophageal cancer. Int J Cancer 2007; 120: 1560-1564.
41. Garavello W, Rossi M, McLaughlin JK et al. Flavonoids and laryngeal cancer risk in Italy. Ann Oncol 2007; 18: 1104-1109.
42. Rossi M, Rosato V, Bosetti C et al. Flavonoids, proanthocyanidins, and the risk of stomach cancer. Cancer Causes Control 2010; 21: 1597-1604.
43. Rossi M, Negri E, Talamini R et al. Flavonoids and colorectal cancer in Italy. Cancer Epidemiol Biomarkers Prev 2006; 15: 1555-1558.
44. Rossi M, Negri E, Parpinel M et al. Proanthocyanidins and the risk of colorectal cancer in Italy. Cancer Causes Control 2010; 21: 243-250.

45. Rossi M, Lugo A, Lagiou P et al. Proanthocyanidins and other flavonoids in relation to pancreatic cancer: a case-control study in Italy. Ann Oncol 2011; Nov 2. [Epub ahead of print].
46. Bosetti C, Spertini L, Parpinel M et al. Flavonoids and breast cancer risk in Italy. Cancer Epidemiol Biomarkers Prev 2005; 14: 805-808.
47. Rossi M, Negri E, Lagiou P et al. Flavonoids and ovarian cancer risk: a case-control study in Italy. Int J Cancer 2008; 123: 895-898.
48. Bosetti C, Bravi F, Talamini R et al. Flavonoids and prostate cancer risk: a study in Italy. Nutr Cancer 2006; 56: 123-127.
49. Bosetti C, Rossi M, McLaughlin JK et al. Flavonoids and the risk of renal cell carcinoma. Cancer Epidemiol Biomarkers Prev 2007; 16: 98-101.
50. Franceschi S, Talamini R, Barra S et al. Smoking and drinking in relation to cancers of the oral cavity, pharynx, larynx, and esophagus in northern Italy. Cancer Res 1990; 50: 6502-6507.
51. La Vecchia C, Chatenoud L, Franceschi S et al. Vegetables and fruit and human cancer: update of an Italian study. Int J Cancer 1999; 82: 151-152.
52. La Vecchia C, Franceschi S. Nutrition and gastric cancer with a focus on Europe. Eur J Cancer Prev 2000; 9: 291-295.
53. Bobe G, Peterson JJ, Gridley G et al. Flavonoid consumption and esophageal cancer among black and white men in the United States. Int J Cancer 2009; 125: 1147-1154.
54. Bors W, Michel C. Antioxidant capacity of flavanols and gallate esters: pulse radiolysis studies. Free Radic Biol Med 1999; 27: 1413-1426.
55. Cotterchio M, Boucher BA, Manno M et al. Dietary phytoestrogen intake is associated with reduced colorectal cancer risk. J Nutr 2006; 136: 3046-3053.
56. Akhter M, Inoue M, Kurahashi N et al. Dietary soy and isoflavone intake and risk of colorectal cancer in the Japan public health center-based prospective study. Cancer Epidemiol Biomarkers Prev 2008; 17: 2128-2135.
57. Oba S, Nagata C, Shimizu N et al. Soy product consumption and the risk of colon cancer: a prospective study in Takayama, Japan. Nutr Cancer 2007; 57: 151-157.
58. Akhter M, Iwasaki M, Yamaji T et al. Dietary isoflavone and the risk of colorectal adenoma: a case-control study in Japan. Br J Cancer 2009; 100: 1812-1816.
59. Theodoratou E, McNeill G, Cetnarskyj R et al. Dietary fatty acids and colorectal cancer: a case-control study. Am J Epidemiol 2007; 166: 181-195.
60. Theodoratou E, Kyle J, Cetnarskyj R et al. Dietary flavonoids and the risk of colorectal cancer. Cancer Epidemiol Biomarkers Prev 2007; 16: 684-693.
61. Cos P, De Bruyne T, Hermans N et al. Proanthocyanidins in health care: current and new trends. Curr Med Chem 2004; 11: 1345-1359.
62. Hoensch HP, Kirch W. Potential role of flavonoids in the prevention of intestinal neoplasia: a review of their mode of action and their clinical perspectives. Int J Gastrointest Cancer 2005; 35: 187-195.
63. Gosse F, Guyot S, Roussi S et al. Chemopreventive properties of apple procyanidins on human colon cancer-derived metastatic SW620 cells and in a rat model of colon carcinogenesis. Carcinogenesis 2005; 26: 1291-1295.
64. Kim YJ, Park HJ, Yoon SH et al. Anticancer effects of oligomeric proanthocyanidins on human colorectal cancer cell line, SNU-C4. World J Gastroenterol 2005; 11: 4674-4678.
65. Kaur M, Singh RP, Gu M et al. Grape seed extract inhibits in vitro and in vivo growth of human colorectal carcinoma cells. Clin Cancer Res 2006; 12: 6194-6202.
66. Faria A, Calhau C, de Freitas V, Mateus N. Procyanidins as antioxidants and tumor cell growth modulators. J Agric Food Chem 2006; 54: 2392-2397.

67. Nandakumar V, Singh T, Katiyar SK. Multi-targeted prevention and therapy of cancer by proanthocyanidins. Cancer Lett 2008; 269: 378-387.
68. Das N, Pereira T. Effects of flavonoids on thermal autoxidation of palm oil – structure activity-relationship. J Am Oil Chem Soc 1990; 67: 255-258.
69. Hayes R, Bookwalter Q, Bayley E. Antioxidant activity of soybean flour and derivatives – a Review. J Food Sci 1977; 42: 1527-1532.
70. Bobe G, Weinstein SJ, Albanes D et al. Flavonoid intake and risk of pancreatic cancer in male smokers (Finland). Cancer Epidemiol Biomarkers Prev 2008; 17: 553-562.
71. Nothlings U, Murphy SP, Wilkens LR et al. Flavonols and pancreatic cancer risk: the multiethnic cohort study. Am J Epidemiol 2007; 166: 924-931.
72. Arts IC, Jacobs DR, Jr., Gross M et al. Dietary catechins and cancer incidence among postmenopausal women: the Iowa Women's Health Study (United States). Cancer Causes Control 2002; 13: 373-382.
73. Knekt P, Jarvinen R, Seppanen R et al. Dietary flavonoids and the risk of lung cancer and other malignant neoplasms. Am J Epidemiol 1997; 146: 223-230.
74. Cotterchio M, Boucher BA, Kreiger N et al. Dietary phytoestrogen intake-lignans and isoflavones-and breast cancer risk (Canada). Cancer Causes Control 2008; 19: 259-272.
75. Zhang M, Xie X, Lee AH, Binns CW. Soy and isoflavone intake are associated with reduced risk of ovarian cancer in southeast china. Nutr Cancer 2004; 49: 125-130.
76. Lukanova A, Kaaks R. Endogenous hormones and ovarian cancer: epidemiology and current hypotheses. Cancer Epidemiol Biomarkers Prev 2005; 14: 98-107.
77. Gercel-Taylor C, Feitelson AK, Taylor DD. Inhibitory effect of genistein and daidzein on ovarian cancer cell growth. Anticancer Res 2004; 24: 795-800.
78. Spinella F, Rosano L, Di Castro V et al. Green tea polyphenol epigallocatechin-3-gallate inhibits the endothelin axis and downstream signaling pathways in ovarian carcinoma. Mol Cancer Ther 2006; 5: 1483-1492.
79. Gates MA, Tworoger SS, Hecht JL et al. A prospective study of dietary flavonoid intake and incidence of epithelial ovarian cancer. Int J Cancer 2007; 121: 2225-2232.
80. Magee PJ, Rowland IR. Phyto-oestrogens, their mechanism of action: current evidence for a role in breast and prostate cancer. Br J Nutr 2004; 91: 513-531.
81. Travis RC, Spencer EA, Allen NE et al. Plasma phyto-oestrogens and prostate cancer in the European Prospective Investigation into Cancer and Nutrition. Br J Cancer 2009; 100: 1817-1823.
82. Wilson RT, Wang J, Chinchilli V et al. Fish, vitamin D, and flavonoids in relation to renal cell cancer among smokers. Am J Epidemiol 2009; 170: 717-729.
83. Pellegrini N, Salvatore S, Valtuena S et al. Development and validation of a food frequency questionnaire for the assessment of dietary total antioxidant capacity. J Nutr 2007; 137: 93-98.
84. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol 2002; 13: 3-9.
85. Nothlings U, Murphy SP, Wilkens LR et al. A food pattern that is predictive of flavonol intake and risk of pancreatic cancer. Am J Clin Nutr 2008; 88: 1653-1662.
86. Edefonti V, Decarli A, La Vecchia C et al. Nutrient dietary patterns and the risk of breast and ovarian cancers. Int J Cancer 2008; 122: 609-613.

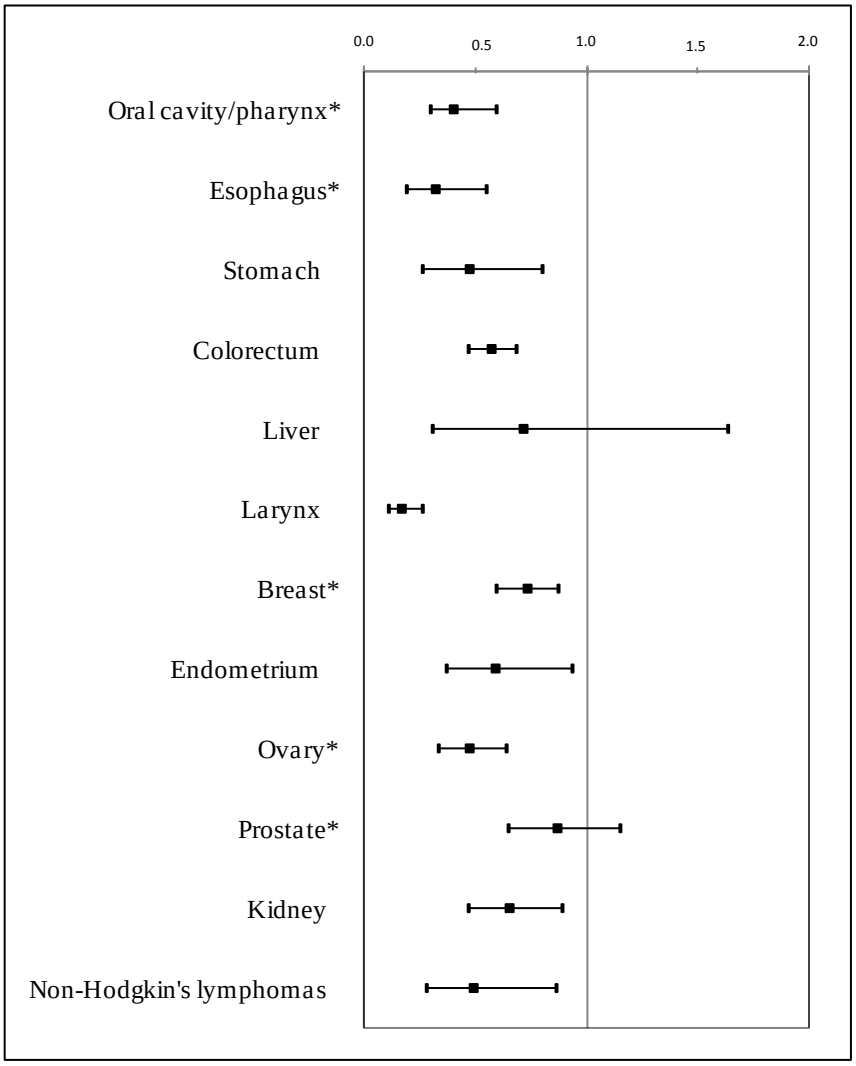
TABLES

Table 1 Primary contributors to flavonoid and proanthocyanidin intake among controls. Italy, 1991-2005.

	Principal flavonoid and proanthocyanidin food sources		
	1	**2**	**3**
Flavonoids			
Flavanols	Tea (50%)	Apples and pears (20%)	Wine (15%)
Flavanones	Citrus fruit (88%)	Fruit juice (11%)	
Flavonols	Apples and pears (16%)	Fennels (16%)	Mixed vegetable salad (12%)
Anthocyanidins	Wine (46%)	Strawberries and cherries (37%)	Onions (7%)
Flavones	Spinaches and Swiss chards (29%)	Vegetable and bean soups (17%)	Tea (15%)
Isoflavones	Soya milk (45%)	Soya (21%)	Vegetable and bean soups (12%)
Total of the above flavonoids	Citrus fruit (28%)	Tea (22%)	Apples and pears (11%)
Proanthocyanidins			
Monomers	Wine (54%)	Apples and pears (18%)	Tea (7%)
Dimers	Wine (50%)	Apples and pears (26%)	Peaches, apricots and prunes (7%)
Trimers	Apples and pears (47%)	Wine (13%)	Peaches, apricots and prunes (10%)
4-6mers	Apples and pears (45%)	Wine (14%)	Vegetable and bean soups (11%)
7-10mers	Apples and pears (48%)	Wine (13%)	Vegetable and bean soups (12%)
> 10 mers	Apples and pears (32%)	Vegetable and bean soups (16%)	Grape (13%)
Total proanthocyanidins	Apples and pears (35%)	Wine (24%)	Vegetable and bean soups (11%)

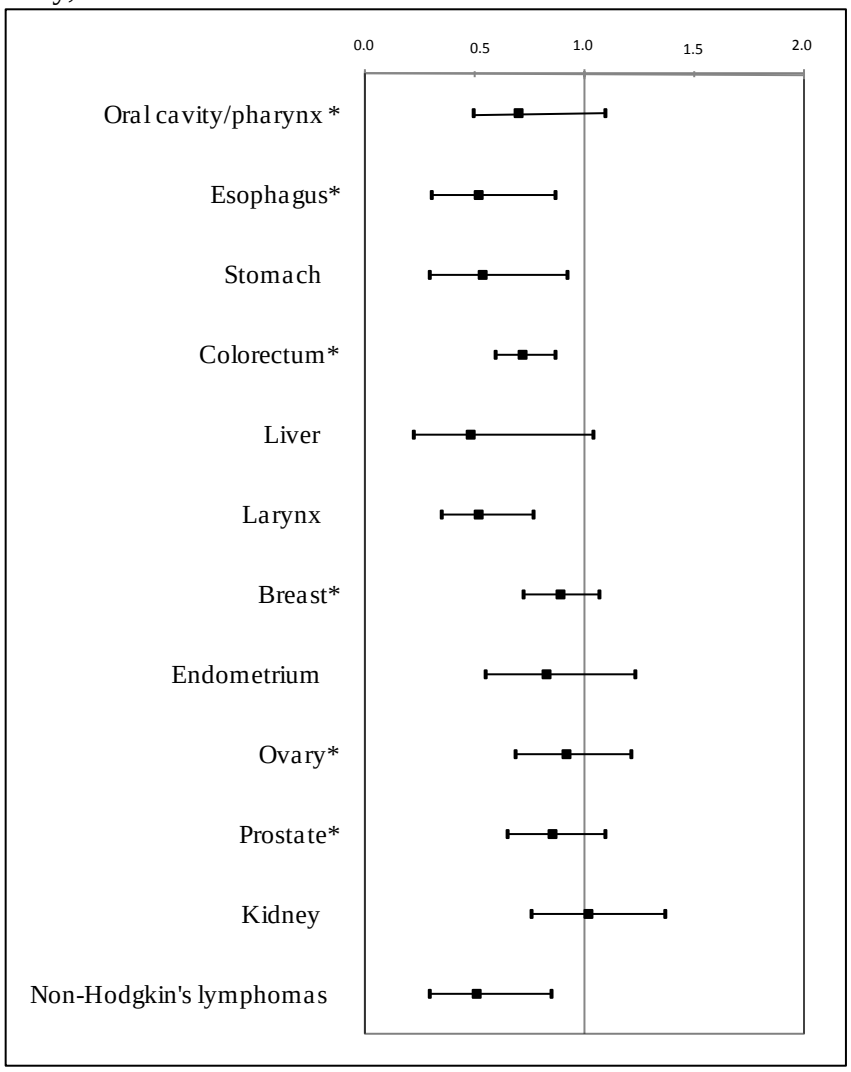
FIGURES

Figure 1 - Odds ratios and 95% confidence intervals of selected cancers for the highest vs the lowest levels of vegetable consumption. Italy, 1991-2005.



* Raw vegetables

Figure 2 - Odds ratios and 95% confidence intervals of selected cancers for the highest vs the lowest levels of fruit consumption. Italy, 1991-2005



* Non-citrus fruit

APPENDIXES

APPENDIX 1 – Chapter 4

Table 1 (Appendix 1) – Multiple logistic regression-derived odds ratios (ORs)[a] and corresponding 95% confidence intervals (CIs) according to intake quintiles of six classes of flavonoids, as well as total flavonoids, among 250 hepatitis B and/or C virus positive hepatocellular carcinoma (HCC) cases, 83 hepatitis B and C virus negative HCC cases, and 6 cholangiocarcinoma cases. Greece, 1995-1998.

Flavonoids	**Quintiles of intake[b]**					**χ^2 trend (p value)**
	1[c]	**2**	**3**	**4**	**5**	
Flavanones						
Upper cutpoint (mg/day)	*57.0*	*103.2*	*109.1*	*120.6*	-	
HCC Virus Positive cases	42	37	50	54	67	1.38 (0.24)
OR	1	0.78	1.10	1.11	1.21	
(95% CI)		(0.43-1.39)	(0.63-1.93)	(0.63-1.94)	(0.69-2.11)	
HCC Virus Negative cases	26	11	9	8	29	0.05 (0.82)
OR	1	0.57	0.42	0.34	1.18	
(95% CI)		(0.24-1.32)	(0.17-1.03)	(0.13-0.85)	(0.56-2.47)	
Cholangiocarcinoma cases[d]	2	1	1	1	1	0.46 (0.50)
Controls	72	72	72	72	72	
Flavan-3-ols						
Upper cutpoint (mg/day)	*25.3*	*40.2*	*51.0*	*66.3*	-	
HCC Virus Positive cases	35	56	39	59	61	0.16 (0.69)
OR	1	1.25	0.87	1.23	1.17	
(95% CI)		(0.70-2.24)	(0.47-1.59)	(0.69-2.19)	(0.65-2.12)	
HCC Virus Negative cases	18	17	14	14	20	0.03 (0.87)
OR	1	1.12	0.87	0.94	1.14	
(95% CI)		(0.48-2.58)	(0.37-2.06)	(0.40-2.25)	(0.50-2.58)	
Cholangiocarcinoma cases[d]	3	2	1	0	0	3.38 (0.066)
Controls	72	72	72	72	72	
Flavonols						
Upper cutpoint (mg/day)	*21.6*	*28.3*	*32.5*	*37.3*	-	
HCC Virus Positive cases	33	54	46	56	61	0.18 (0.67)
OR	1	1.75	1.21	1.41	1.35	
(95% CI)		(0.98-3.12)	(0.66-2.21)	(0.78-2.53)	(0.75-2.44)	
HCC Virus Negative cases	20	17	13	12	21	0.02 (0.89)
OR	1	0.82	0.58	0.63	1.11	
(95% CI)		(0.37-1.81)	(0.25-1.37)	(0.27-1.51)	(0.51-2.43)	
Cholangiocarcinoma cases[d]	1	4	0	1	0	1.80 (0.18)
Controls	72	72	72	72	72	
Anthocyanidins						
Upper cutpoint (mg/day)	*10.2*	*40.7*	*64.6*	*152.7*	-	
HCC Virus Positive cases	35	70	33	41	71	0.16 (0.68)
OR	1	1.31	0.97	0.76	1.42	
(95% CI)		(0.77-2.25)	(0.51-1.85)	(0.41-1.41)	(0.81-2.47)	
HCC Virus Negative cases	13	24	9	15	22	0.49 (0.48)
OR	1	1.82	0.97	1.19	1.73	
(95% CI)		(0.81-4.09)	(0.33-2.81)	(0.47-3.00)	(0.74-4.04)	
Cholangiocarcinoma cases[d]	3	2	1	0	0	3.64 (0.056)
Controls	72	93	51	66	78	

	Quintiles of intake[b]					χ^2 trend
	1[b]	**2**	**3**	**4**	**5**	**(p value)**
Flavones						
Upper cutpoint (mg/day)	*0.25*	*0.58*	*0.90*	*1.16*	-	
HCC Virus Positive cases	44	49	54	64	40	3.87 (0.049)
OR	1	0.85	0.84	0.84	0.50	
(95% CI)		(0.48-1.51)	(0.46-1.51)	(0.47-1.52)	(0.27-0.94)	
HCC Virus Negative cases	19	22	12	20	10	2.99 (0.084)
OR	1	1.03	0.52	0.91	0.41	
(95% CI)		(0.48-2.19)	(0.21-1.26)	(0.40-2.08)	(0.16-1.06)	
Cholangiocarcinoma cases[d]	0	2	0	1	3	1.41 (0.24)
Controls	72	72	71	71	74	
Isoflavones						
Upper cutpoint (mg/day)	*0.03*	*0.04*	*0.07*	*0.32*	-	
HCC Virus Positive cases	38	42	49	54	67	1.36 (0.24)
OR	1	0.96	0.97	0.83	0.74	
(95% CI)		(0.55-1.67)	(0.53-1.47)	(0.48-1.46)	(0.41-1.32)	
HCC Virus Negative cases	20	13	12	19	19	0.02 (0.88)
OR	1	0.53	0.47	0.88	0.85	
(95% CI)		(0.22-1.24)	(0.19-1.17)	(0.41-1.89)	(0.39-1.86)	
Cholangiocarcinoma cases[d]	3	1	0	1	1	1.51 (0.22)
Controls	72	72	72	72	72	
Total flavonoids						
Upper cutpoint (mg/day)	*145.8*	*212.6*	*256.8*	*358.1*	-	
HCC Virus Positive cases	38	42	47	56	67	0.68 (0.41)
OR	1	0.93	0.91	0.98	1.22	
(95% CI)		(0.52-1.66)	(0.51-1.65)	(0.54-1.77)	(0.69-2.16)	
HCC Virus Negative cases	22	14	6	19	22	0.09 (0.76)
OR	1	0.74	0.28	0.81	0.96	
(95% CI)		(0.33-1.70)	(0.10-0.82)	(0.35-1.84)	(0.43-2.12)	
Cholangiocarcinoma cases[d]	3	1	2	0	0	3.61 (0.057)
Controls	72	72	72	72	72	

[a]Adjusted for gender, age, education, tobacco smoking, and total energy intake. [b]Control generated quintiles [c]Reference category. [d]Adjusted only for gender; quintile specific odds ratios not calculated because of sparsity of data.

Figure 1 (Appendix 1) - Odds ratios and 95% confidence intervals of cancers at the upper aerodigestive tract for the highest versus the lowest quintile of intake of selected classes of flavonoids. Italy, 1992-2005.

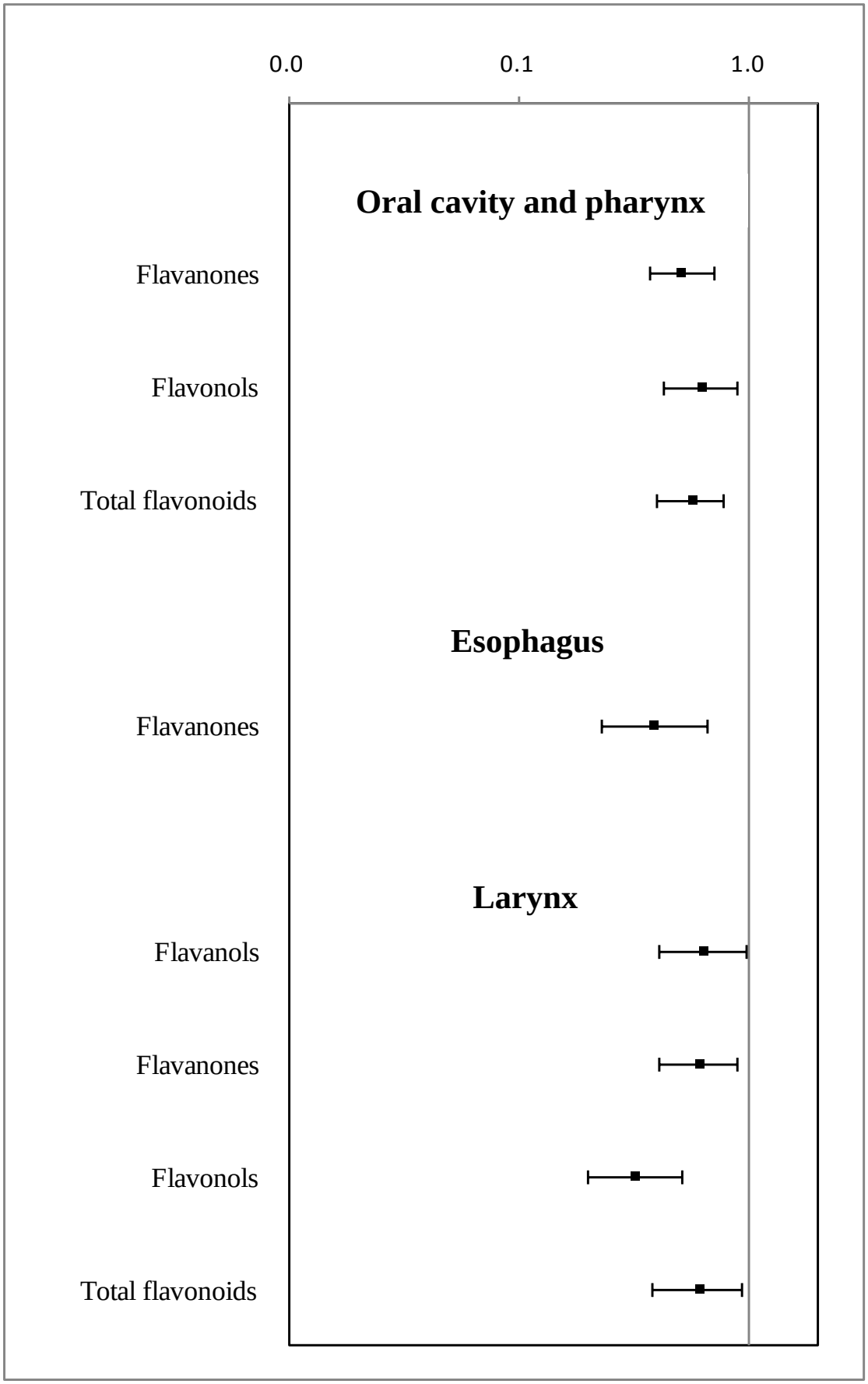


Figure 2 (Appendix 1) - Odds ratios and 95% confidence intervals of stomach cancer for the highest versus the lowest quintile of intake of six classes of flavonoids and two classes of proanthocyanidins. Italy, 1997-2007.

.

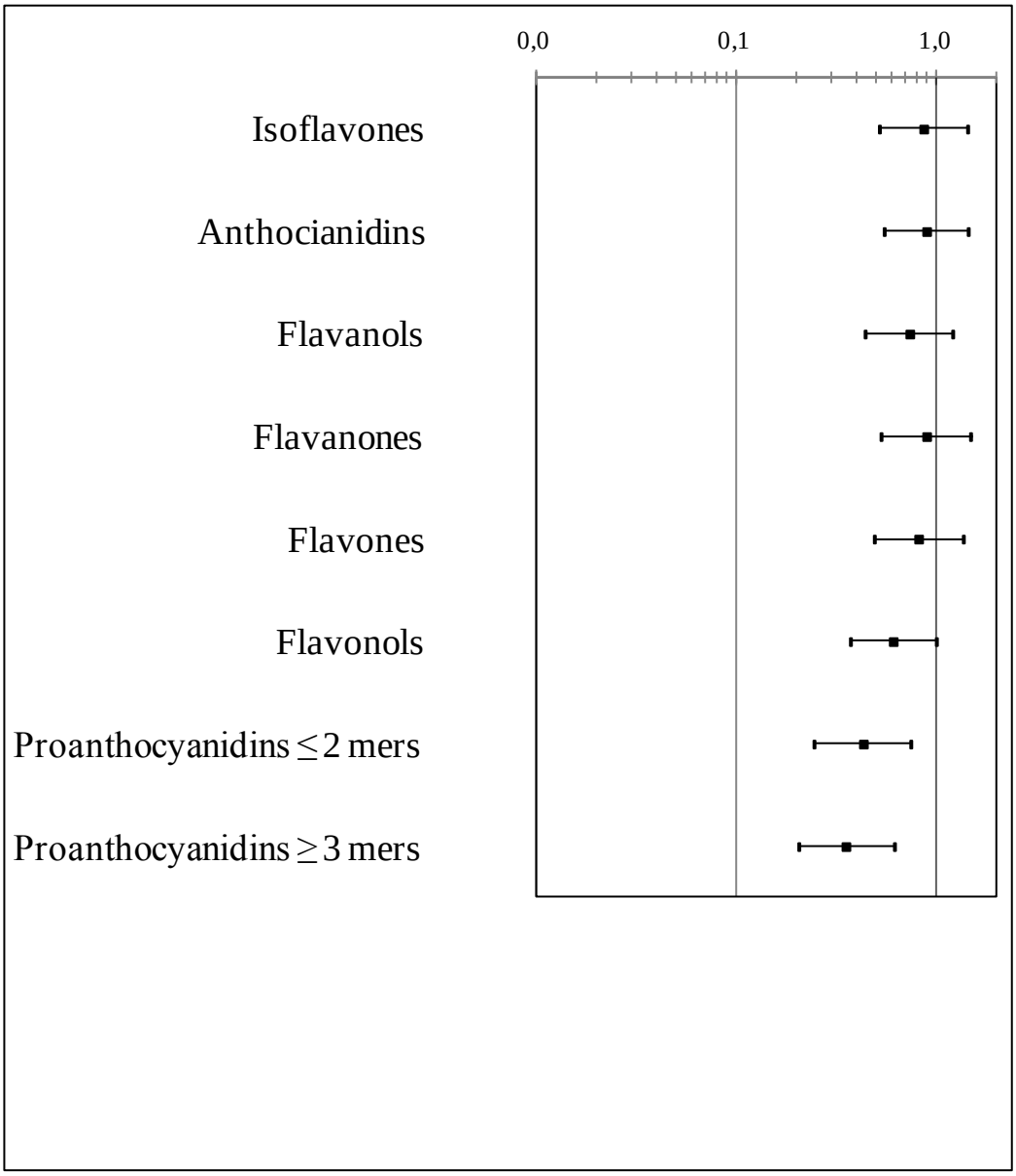


Figure 3 (Appendix 1) - Odds ratios and 95% confidence intervals of colorectal cancer for the highest versus the lowest quintile of intake of selected classes of flavonoids and proanthocyanidins. Italy, 1992-1996.

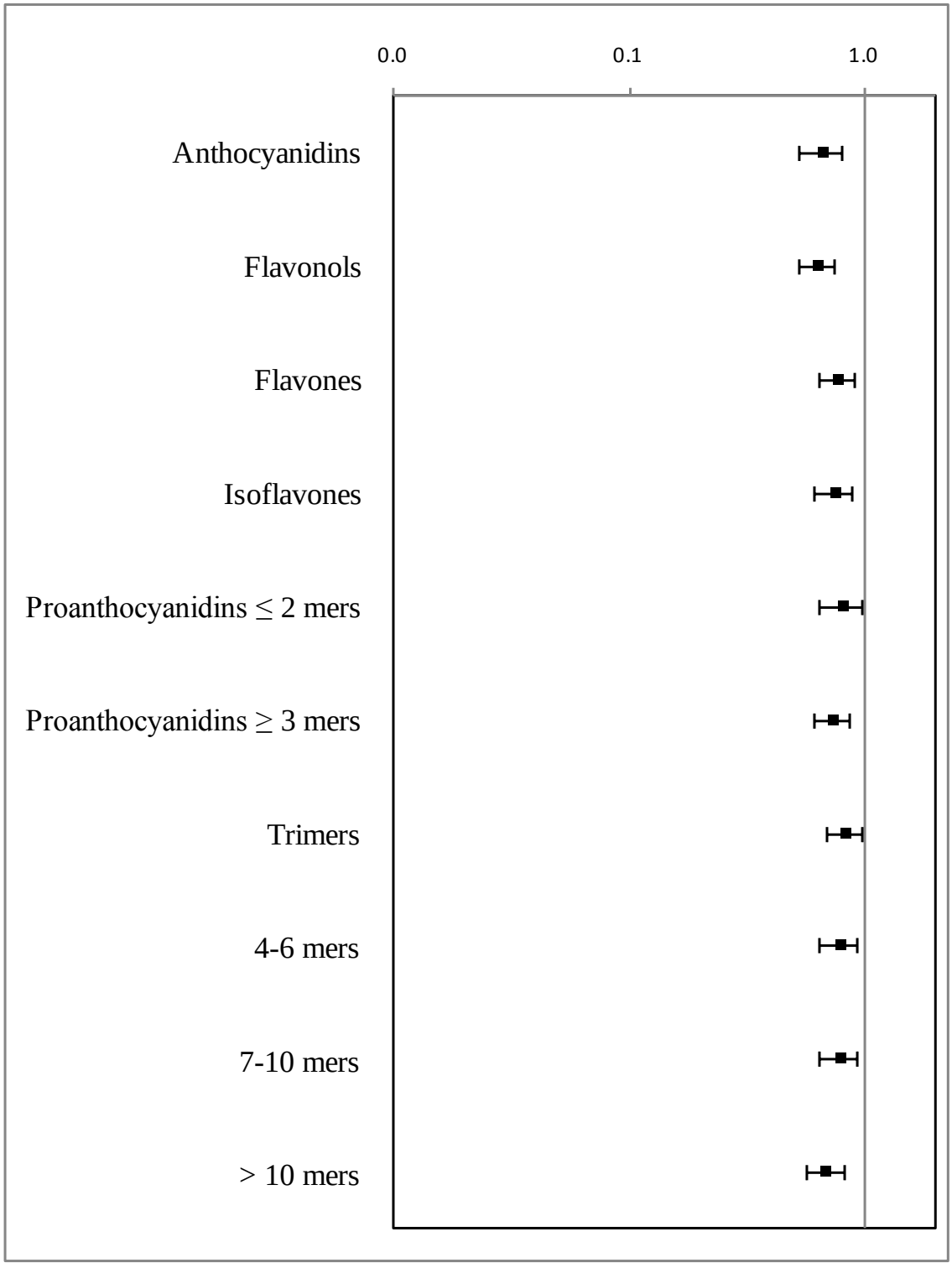


Figure 4 (Appendix 1) - Odds ratios and 95% confidence intervals of pancreatic cancer for the highest versus the lowest quintile of intake of selected classes of flavonoids and proanthocyanidins. Italy, 1991-2008.

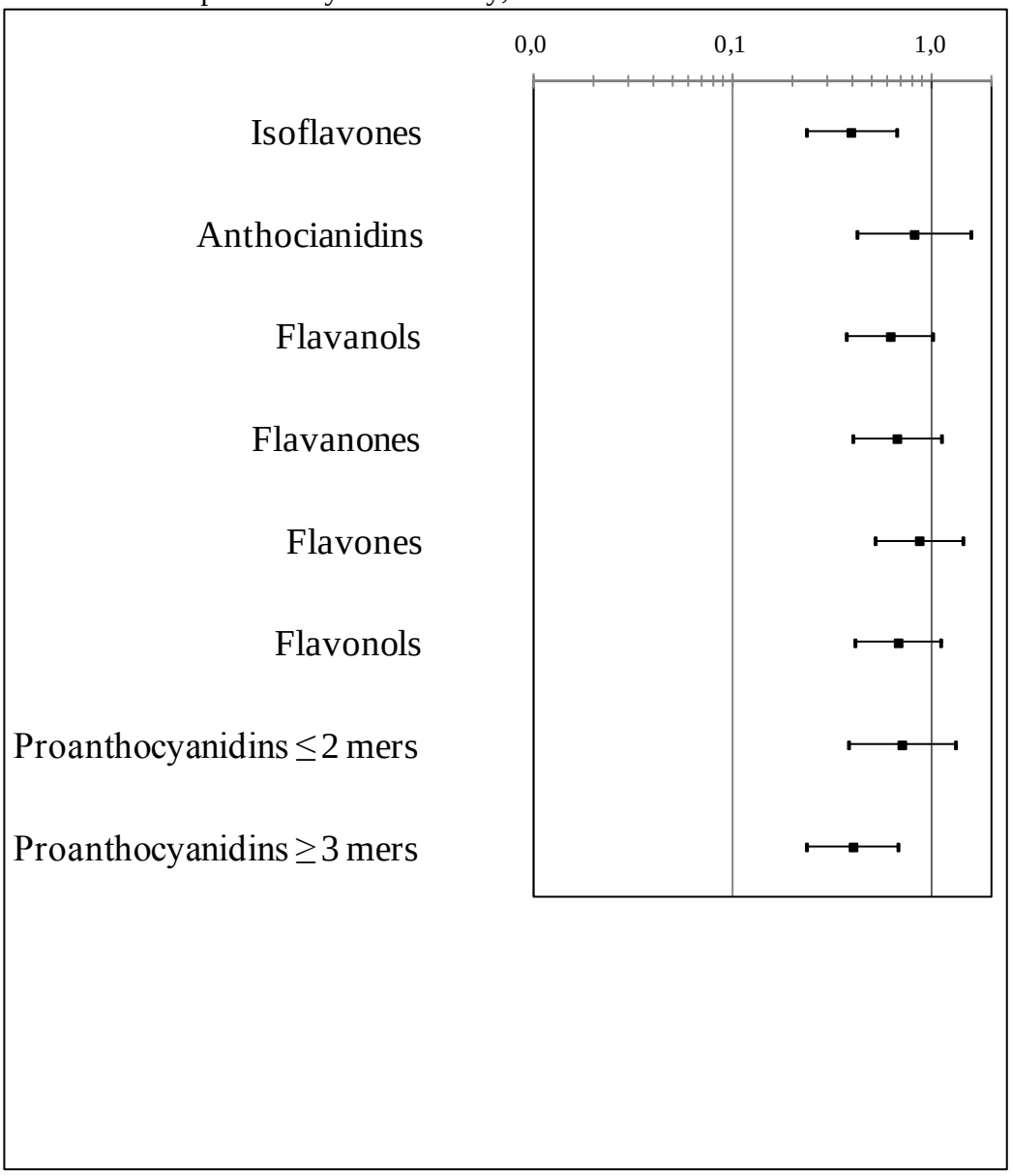
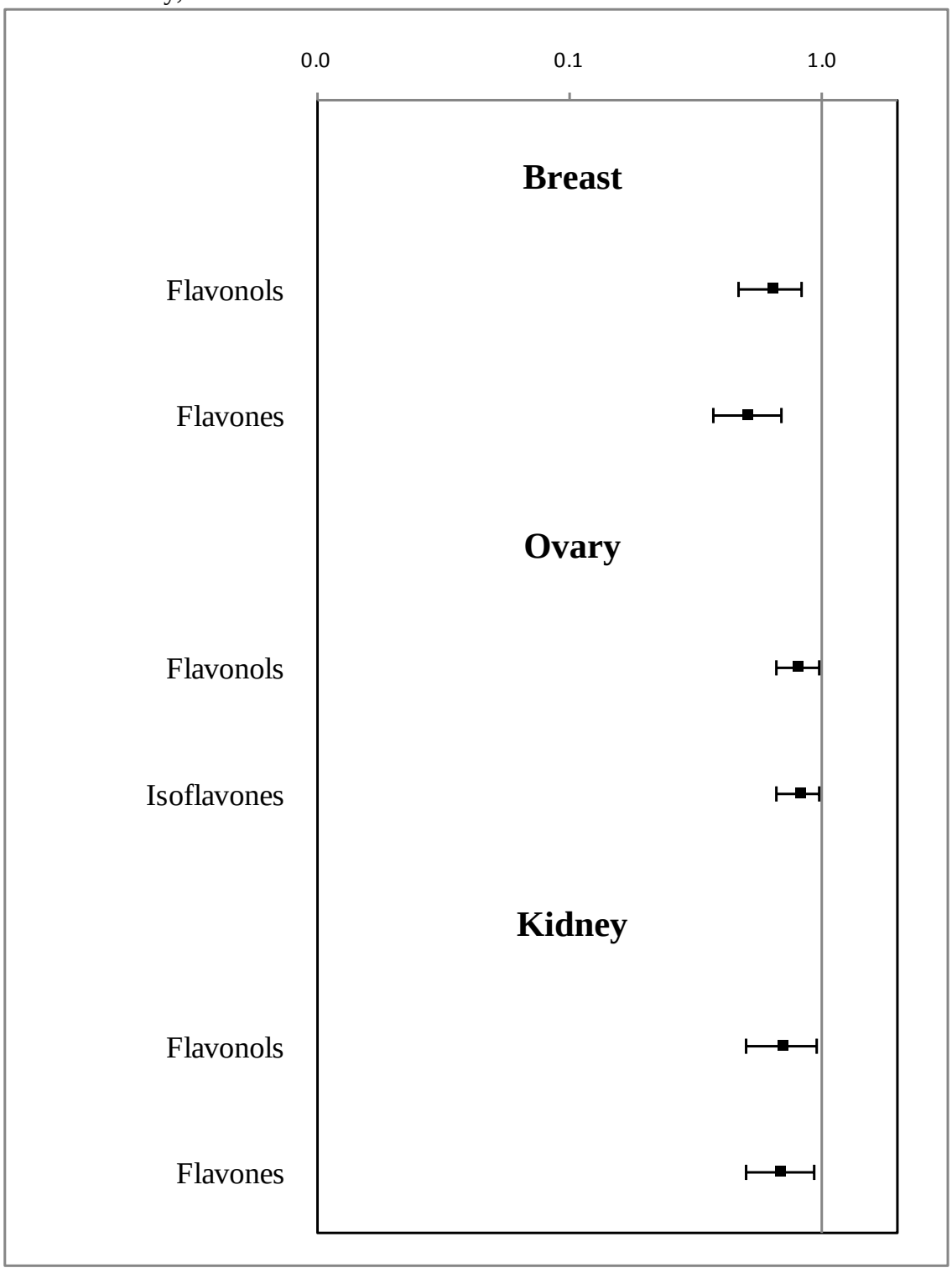


Figure 5 (Appendix 1) - Odds ratios and 95% confidence intervals of breast, ovary, and kidney cancers for the highest versus the lowest quintile of intake of selected classes of flavonoids. Italy, 1991-2004.



APPENDIX 2 – Chapter 5

Table 1 (Appendix 2) - Distribution of 1,953 cases with colorectal cancer and 4,154 controls according to quintiles of energy-adjusted classes of TAC among controls. (Italy, 1992-1996)

	Mean (SD)[a]	Quintiles of intake[b]				
		1	2	3	4	5
TAC						
TEAC (mmol/d)	4.47 (2.55)					
Upper cutoff point		3.17	3.93	4.59	5.54	-
Cases		389	397	362	389	416
TRAP (mmol/d)	4.56 (3.09)					
Upper cutoff point		2.88	3.78	4.61	5.94	-
Cases		398	364	376	392	423
FRAP (mmol/d)	11.45 (6.65)					
Upper cutoff point		7.93	9.93	11.75	14.34	-
Cases		393	375	379	384	422

[a] Mean intake and standard deviation (SD) among control distribution.

[b] Based on the control distribution. Number of controls across quintiles could slightly differ between 828 and 831 according to the distributions of each measure of TAC.

Table 2 (Appendix 2) - Correlation of TEAC, TRAP and FRAP with selected covariates among 4,154 controls. (Italy, 1992-1996)

	TEAC	TRAP	FRAP
Fruit	0.29	0.19	0.27
Vegetables	0.24	0.16	0.21
Flavonoids	0.64	0.56	0.65
Anthocyanidins	0.86	0.89	0.86
Flavones	0.04	-0.02	0.05
Flavonols	0.30	0.27	0.30
Proanthocyanidins	0.62	0.59	0.59
Polymers ≥ 10 mers	0.35	0.30	0.32
Vitamin C	0.35	0.21	0.33
Carotene	0.21	0.13	0.19
Vitamin E	0.39	0.25	0.32
Vitamin D	0.17	0.10	0.14
Beta carotene	0.23	0.14	0.20
Energy	0.63	0.53	0.59

Table 3 (Appendix 2) - Odds ratios (ORs) [a] of colorectal cancer among 1,953 cases with colorectal cancer and 4,154 controls, and corresponding 95% confidence intervals (CIs) according to quintiles[b] (I-V) of three energy-adjusted total antioxidant capacity indexes. (Italy, 1992-1996)

	OR (95% CI)					χ^2 trend (p-value)	OR continuous [d]
	Quintiles [b]						
	I[c]	II	III	IV	V		
TEAC	-	0.95 (0.79-1.13)	0.78 (0.65-0.94)	0.77 (0.64-0.93)	0.78 (0.63-0.96)	0.002	0.89 (0.82-0.97)
TRAP	-	0.83 (0.70-1.00)	0.77 (0.64-0.92)	0.70 (0.58-0.85)	0.74 (0.60-0.92)	0.001	0.90 (0.83-0.98)
FRAP	-	0.87 (0.73-1.05)	0.79 (0.66-0.95)	0.74 (0.62-0.90)	0.78 (0.63-0.96)	0.003	0.89 (0.82-0.98)

[a] Estimated using multiple logistic regression models adjusted for sex, age, study centre, family history, education, alcohol consumption, body mass index, physical activity and energy intake, according to the residual model. [b] Based on the control distribution. [c] Reference category. [d] Estimated for an increment of intake equal to the difference between the upper cut-off points of the 4th and the 1st quintiles.

APPENDIX 3 – Chapter 6

Table 1 (Appendix 3) - Distribution of 34 flavonoids in 1953 cases and 4,154 controls from the study on colorectal cancer. Italy, 1992-1996.

Variables		Median	St dev	Minimum	Maximum
FL1	ISO-DAIDZEINA	30.34	221.74	0.00	10037.06
FL2	ISO-GENISTEIN	33.00	301.05	0.21	13663.07
FL3	ANTHO-Cyanidin	6.83	6.65	0.01	117.15
FL4	ANTHO-Delphinidin	1.00	1.30	0.00	17.14
FL5	ANTHO-Malvidin	8.38	10.90	0.00	144.05
FL6	ANTHO-Pelargonidin	0.03	0.05	0.00	1.03
FL7	ANTHO-Peonidin	1.91	2.23	0.00	29.49
FL8	ANTHO-Petunidin	2.04	2.65	0.00	35.03
FL9	FLAVAN3-(+)-Catechin	17.38	15.64	0.00	205.57
FL10	FLAVAN3-(-)-Epigallocatechin	2.96	6.30	0.00	106.92
FL11	FLAVAN3-(-)-Epicatechin	20.45	12.42	0.00	147.25
FL12	FLAVAN3-(-)-Epicatechin 3-gallate	2.46	4.29	0.00	72.22
FL13	FLAVAN3-(-)-Epigallocatechin 3-gallate	6.81	14.52	0.00	246.38
FL14	FLAVAN3-Theaflavin	0.10	0.22	0.00	3.78
FL15	FLAVAN3-Thearubigins	8.48	18.08	0.00	306.72
FL16	FLAVA-Eriodictyol	0.07	0.12	0.00	1.78
FL17	FLAVA-Hesperetin	24.35	20.90	0.00	230.85
FL18	FLAVA-naringenin	13.71	12.37	0.00	178.86
FL19	FLAVONE-Apigenin	0.23	0.14	0.00	1.68
FL20	FLAVONE-Luteolin	0.25	0.19	0.00	2.69
FL21	FLAVONO-Isorhamnetin	1.47	1.29	0.00	19.48
FL22	FLAVONO-Kaempferol	1.65	1.07	0.05	15.65
FL23	FLAVONO-Myricetin	2.39	2.31	0.01	39.37
FL24	FLAVONO-Quercetin	15.99	7.27	1.07	109.12
FL25	FLAVAN3-Theaflavin-3,3'-digallate	0.11	0.23	0.00	3.92
FL26	FLAVAN3-Theaflavin-3'-gallate	0.08	0.18	0.00	2.97
FL27	FLAVAN3-Theaflavin-3-gallate	0.09	0.18	0.00	3.11
FL28	FLAVAN3-(+)-Gallocatechin	0.47	1.00	0.00	17.01
FL29	Proanthocyanidin-Monomers	44.68	35.16	0.04	430.27
FL30	Proanthocyanidin-Dimers	57.74	42.60	0.00	522.53
FL31	Proanthocyanidin-Trimers	19.64	11.46	0.00	157.57
FL32	Proanthocyanidin-4-6mers	67.04	36.36	0.00	459.57
FL33	Proanthocyanidin-7-10mers	53.82	28.93	0.00	387.76
FL34	Proanthocyanidin-Polymers	115.93	60.02	0.00	675.39

Table 2 (Appendix 3) - Pearson correlation coefficients and Prob>|r| with H0: Rho=0 among 34 flavonoids in 1953 cases and 4,154 controls from the study on colorectal cancer. Italy, 1992-1996.

	FL1	**FL2**	**FL3**	**FL4**	**FL5**	**FL6**	**FL7**	**FL8**	**FL9**
FL1 ISO-DAIDZEINA	1.00000	0.99821 <.0001	0.00987 0.4407	-0.01774 0.1657	-0.01774 0.1657	0.01004 0.4328	-0.01636 0.2013	-0.01774 0.1657	-0.01380 0.2811
FL2 ISO-GENISTEIN	0.99821 <.0001	1.00000	0.01053 0.4108	-0.01404 0.2727	-0.01404 0.2728	0.00946 0.4597	-0.01273 0.3199	-0.01404 0.2728	-0.01449 0.2575
FL3 ANTHO-Cyanidin	0.00987 0.4407	0.01053 0.4108	1.00000	0.04896 0.0001	0.04892 0.0001	0.92414 <.0001	0.17070 <.0001	0.04889 0.0001	0.12474 <.0001
FL4 ANTHO-Delphinidin	-0.01774 0.1657	-0.01404 0.2727	0.04896 0.0001	1.00000	1.00000 <.0001	-0.05597 <.0001	0.99179 <.0001	1.00000 <.0001	0.97494 <.0001
FL5 ANTHO-Malvidin	-0.01774 0.1657	-0.01404 0.2728	0.04892 0.0001	1.00000 <.0001	1.00000	-0.05624 <.0001	0.99178 <.0001	1.00000 <.0001	0.97495 <.0001
FL6 ANTHO-Pelargonidin	0.01004 0.4328	0.00946 0.4597	0.92414 <.0001	-0.05597 <.0001	-0.05624 <.0001	1.00000	0.06733 <.0001	-0.05636 <.0001	0.01658 0.1951
FL7 ANTHO-Peonidin	-0.01636 0.2013	-0.01273 0.3199	0.17070 <.0001	0.99179 <.0001	0.99178 <.0001	0.06733 <.0001	1.00000	0.99178 <.0001	0.97643 <.0001
FL8 ANTHO-Petunidin	-0.01774 0.1657	-0.01404 0.2728	0.04889 0.0001	1.00000 <.0001	1.00000 <.0001	-0.05636 <.0001	0.99178 <.0001	1.00000	0.97494 <.0001
FL9 FLAVAN3-(+)-Catechin	-0.01380 0.2811	-0.01449 0.2575	0.12474 <.0001	0.97494 <.0001	0.97495 <.0001	0.01658 0.1951	0.97643 <.0001	0.97494 <.0001	1.00000
FL10 FLAVAN3-(-)-Epigallocatechin	-0.00846 0.5084	-0.00859 0.5024	-0.00547 0.6691	-0.02112 0.0989	-0.02110 0.0993	-0.01077 0.4000	-0.02226 0.0819	-0.02111 0.0990	0.09507 <.0001
FL11 FLAVAN3-(-)-Epicatechin	0.02416 0.0590	-0.00998 0.4355	0.17431 <.0001	0.66830 <.0001	0.66831 <.0001	0.08882 <.0001	0.67962 <.0001	0.66829 <.0001	0.77095 <.0001
FL12 FLAVAN3-(-)-Epicatechin 3-gallate	-0.00808 0.5279	-0.00845 0.5091	0.02761 0.0310	-0.02191 0.0868	-0.02189 0.0872	0.01985 0.1210	-0.01900 0.1377	-0.02190 0.0870	0.11224 <.0001
FL13 FLAVAN3-(-)-Epigallocatechin 3-gallate	-0.00846 0.5084	-0.00859 0.5023	-0.00548 0.6686	-0.02112 0.0988	-0.02110 0.0992	-0.01078 0.3996	-0.02227 0.0819	-0.02111 0.0990	0.09506 <.0001
FL14 FLAVAN3-Theaflavin	-0.00846 0.5084	-0.00859 0.5023	-0.00548 0.6686	-0.02112 0.0988	-0.02110 0.0992	-0.01078 0.3996	-0.02227 0.0819	-0.02111 0.0990	0.09506 <.0001
FL15 FLAVAN3-Thearubigins	-0.00846 0.5084	-0.00859 0.5023	-0.00548 0.6686	-0.02112 0.0988	-0.02110 0.0992	-0.01078 0.3996	-0.02227 0.0819	-0.02111 0.0990	0.09506 <.0001
FL16 FLAVA-Eriodictyol	0.02998 0.0191	0.02994 0.0193	0.11961 <.0001	-0.05453 <.0001	-0.05452 <.0001	0.10726 <.0001	-0.04032 0.0016	-0.05454 <.0001	-0.01297 0.3107
FL17 FLAVA-Hesperetin	-0.00028 0.9828	-0.00631 0.6223	0.21014 <.0001	-0.06289 <.0001	-0.06289 <.0001	0.20325 <.0001	-0.03615 0.0047	-0.06291 <.0001	0.02238 0.0803
FL18 FLAVA-naringenin	0.01093 0.3933	0.00583 0.6489	0.19985 <.0001	-0.07837 <.0001	-0.07837 <.0001	0.19021 <.0001	-0.05326 <.0001	-0.07839 <.0001	-0.00271 0.8325
FL19 FLAVONE-Apigenin	-0.00661 0.6056	-0.00194 0.8797	0.16096 <.0001	0.10143 <.0001	0.10146 <.0001	0.05472 <.0001	0.10896 <.0001	0.10144 <.0001	0.18733 <.0001

FL20 FLAVONE-Luteolin	0.00087 0.9459	-0.00040 0.9749	0.12492 <.0001	0.00059 0.9633	0.00061 0.9619	0.08293 <.0001	0.01134 0.3758	0.00058 0.9641	0.07687 <.0001
FL21 FLAVONO-Isorhamnetin	0.00730 0.5685	0.00857 0.5030	0.31243 <.0001	0.09379 <.0001	0.09379 <.0001	0.08206 <.0001	0.10429 <.0001	0.09377 <.0001	0.12383 <.0001
FL22 FLAVONO-Kaempferol	0.00600 0.6394	0.00421 0.7425	0.16465 <.0001	0.11305 <.0001	0.11308 <.0001	0.08925 <.0001	0.12486 <.0001	0.11305 <.0001	0.21774 <.0001
FL23 FLAVONO-Myricetin	-0.00650 0.6114	-0.00557 0.6632	0.08360 <.0001	0.60236 <.0001	0.60238 <.0001	0.00971 0.4483	0.60330 <.0001	0.60236 <.0001	0.63963 <.0001
FL24 FLAVONO-Quercetin	0.02933 0.0219	0.00847 0.5082	0.26715 <.0001	0.26833 <.0001	0.26834 <.0001	0.12738 <.0001	0.28493 <.0001	0.26832 <.0001	0.37464 <.0001
FL25 FLAVAN3-Theaflavin-3,3'-digallate	-0.00846 0.5084	-0.00859 0.5023	-0.00548 0.6686	-0.02112 0.0988	-0.02110 0.0992	-0.01078 0.3996	-0.02227 0.0819	-0.02111 0.0990	0.09506 <.0001
FL26 FLAVAN3-Theaflavin-3'-gallate	-0.00846 0.5084	-0.00859 0.5023	-0.00548 0.6686	-0.02112 0.0988	-0.02110 0.0992	-0.01078 0.3996	-0.02227 0.0819	-0.02111 0.0990	0.09506 <.0001
FL27 FLAVAN3-Theaflavin-3-gallate	-0.00846 0.5084	-0.00859 0.5023	-0.00548 0.6686	-0.02112 0.0988	-0.02110 0.0992	-0.01078 0.3996	-0.02227 0.0819	-0.02111 0.0990	0.09506 <.0001
FL28 FLAVAN3-(+)-Gallocatechin	-0.00846 0.5084	-0.00859 0.5023	-0.00548 0.6686	-0.02112 0.0988	-0.02110 0.0992	-0.01078 0.3996	-0.02227 0.0819	-0.02111 0.0990	0.09506 <.0001
FL29 Proanthocyanidin-Monomers	-0.00530 0.6785	-0.01185 0.3546	0.12265 <.0001	0.90762 <.0001	0.90763 <.0001	0.01652 0.1968	0.90909 <.0001	0.90762 <.0001	0.94713 <.0001
FL30 Proanthocyanidin-Dimers	0.00155 0.9035	-0.01073 0.4019	0.13388 <.0001	0.90819 <.0001	0.90819 <.0001	0.02627 0.0401	0.91092 <.0001	0.90818 <.0001	0.94521 <.0001
FL31 Proanthocyanidin-Trimers	0.04046 0.0016	0.00516 0.6871	0.23331 <.0001	0.26243 <.0001	0.26244 <.0001	0.17204 <.0001	0.28491 <.0001	0.26241 <.0001	0.37760 <.0001
FL32 Proanthocyanidin-4-6mers	0.04170 0.0011	0.00594 0.6428	0.26208 <.0001	0.31610 <.0001	0.31612 <.0001	0.18929 <.0001	0.34091 <.0001	0.31609 <.0001	0.43394 <.0001
FL33 Proanthocyanidin-7-10mers	0.04539 0.0004	0.00701 0.5838	0.25419 <.0001	0.29333 <.0001	0.29334 <.0001	0.17983 <.0001	0.31698 <.0001	0.29332 <.0001	0.41364 <.0001
FL34 Proanthocyanidin-Polymers	0.03587 0.0051	0.01013 0.4286	0.33152 <.0001	0.32649 <.0001	0.32651 <.0001	0.24344 <.0001	0.35854 <.0001	0.32648 <.0001	0.45822 <.0001

	FL10	**FL11**	**FL12**	**FL13**	**FL14**	**FL15**	**FL16**	**FL17**	**FL18**
FL1 ISO-DAIDZEINA	-0.00846 0.5084	0.02416 0.0590	-0.00808 0.5279	-0.00846 0.5084	-0.00846 0.5084	-0.00846 0.5084	0.02998 0.0191	-0.00028 0.9828	0.01093 0.3933
FL2 ISO-GENISTEIN	-0.00859 0.5024	-0.00998 0.4355	-0.00845 0.5091	-0.00859 0.5023	-0.00859 0.5023	-0.00859 0.5023	0.02994 0.0193	-0.00631 0.6223	0.00583 0.6489
FL3 ANTHO-Cyanidin	-0.00547 0.6691	0.17431 <.0001	0.02761 0.0310	-0.00548 0.6686	-0.00548 0.6686	-0.00548 0.6686	0.11961 <.0001	0.21014 <.0001	0.19985 <.0001
FL4 ANTHO-Delphinidin	-0.02112 0.0989	0.66830 <.0001	-0.02191 0.0868	-0.02112 0.0988	-0.02112 0.0988	-0.02112 0.0988	-0.05453 <.0001	-0.06289 <.0001	-0.07837 <.0001

FL5 ANTHO-Malvidin	-0.02110 0.0993	0.66831 <.0001	-0.02189 0.0872	-0.02110 0.0992	-0.02110 0.0992	-0.02110 0.0992	-0.05452 <.0001	-0.06289 <.0001	-0.07837 <.0001
FL6 ANTHO-Pelargonidin	-0.01077 0.4000	0.08882 <.0001	0.01985 0.1210	-0.01078 0.3996	-0.01078 0.3996	-0.01078 0.3996	0.10726 <.0001	0.20325 <.0001	0.19021 <.0001
FL7 ANTHO-Peonidin	-0.02226 0.0819	0.67962 <.0001	-0.01900 0.1377	-0.02227 0.0819	-0.02227 0.0819	-0.02227 0.0819	-0.04032 0.0016	-0.03615 0.0047	-0.05326 <.0001
FL8 ANTHO-Petunidin	-0.02111 0.0990	0.66829 <.0001	-0.02190 0.0870	-0.02111 0.0990	-0.02111 0.0990	-0.02111 0.0990	-0.05454 <.0001	-0.06291 <.0001	-0.07839 <.0001
FL9 FLAVAN3-(+)-Catechin	0.09507 <.0001	0.77095 <.0001	0.11224 <.0001	0.09506 <.0001	0.09506 <.0001	0.09506 <.0001	-0.01297 0.3107	0.02238 0.0803	-0.00271 0.8325
FL10 FLAVAN3-(-)-Epigallocatechin	1.00000	0.22078 <.0001	0.99212 <.0001	1.00000 <.0001	1.00000 <.0001	1.00000 <.0001	0.09961 <.0001	0.01517 0.2360	0.01932 0.1312
FL11 FLAVAN3-(-)-Epicatechin	0.22078 <.0001	1.00000	0.24365 <.0001	0.22076 <.0001	0.22076 <.0001	0.22076 <.0001	0.05629 <.0001	0.12828 <.0001	0.10041 <.0001
FL12 FLAVAN3-(-)-Epicatechin 3-gallate	0.99212 <.0001	0.24365 <.0001	1.00000	0.99212 <.0001	0.99212 <.0001	0.99212 <.0001	0.11019 <.0001	0.04705 0.0002	0.04721 0.0002
FL13 FLAVAN3-(-)-Epigallocatechin 3-gallate	1.00000 <.0001	0.22076 <.0001	0.99212 <.0001	1.00000	1.00000 <.0001	1.00000 <.0001	0.09964 <.0001	0.01519 0.2354	0.01934 0.1307
FL14 FLAVAN3-Theaflavin	1.00000 <.0001	0.22076 <.0001	0.99212 <.0001	1.00000 <.0001	1.00000	1.00000 <.0001	0.09964 <.0001	0.01519 0.2354	0.01934 0.1307
FL15 FLAVAN3-Thearubigins	1.00000 <.0001	0.22076 <.0001	0.99212 <.0001	1.00000 <.0001	1.00000 <.0001	1.00000	0.09964 <.0001	0.01519 0.2354	0.01934 0.1307
FL16 FLAVA-Eriodictyol	0.09961 <.0001	0.05629 <.0001	0.11019 <.0001	0.09964 <.0001	0.09964 <.0001	0.09964 <.0001	1.00000	0.15487 <.0001	0.46482 <.0001
FL17 FLAVA-Hesperetin	0.01517 0.2360	0.12828 <.0001	0.04705 0.0002	0.01519 0.2354	0.01519 0.2354	0.01519 0.2354	0.15487 <.0001	1.00000	0.86545 <.0001
FL18 FLAVA-naringenin	0.01932 0.1312	0.10041 <.0001	0.04721 0.0002	0.01934 0.1307	0.01934 0.1307	0.01934 0.1307	0.46482 <.0001	0.86545 <.0001	1.00000
FL19 FLAVONE-Apigenin	0.52752 <.0001	0.25554 <.0001	0.53620 <.0001	0.52753 <.0001	0.52753 <.0001	0.52753 <.0001	0.08416 <.0001	0.05154 <.0001	0.04952 0.0001
FL20 FLAVONE-Luteolin	0.39969 <.0001	0.19456 <.0001	0.40653 <.0001	0.39969 <.0001	0.39969 <.0001	0.39969 <.0001	0.12548 <.0001	0.08212 <.0001	0.09818 <.0001
FL21 FLAVONO-Isorhamnetin	0.03826 0.0028	0.15271 <.0001	0.04772 0.0002	0.03827 0.0028	0.03827 0.0028	0.03827 0.0028	0.10907 <.0001	0.08451 <.0001	0.10435 <.0001
FL22 FLAVONO-Kaempferol	0.67437 <.0001	0.33159 <.0001	0.67674 <.0001	0.67439 <.0001	0.67439 <.0001	0.67439 <.0001	0.14987 <.0001	0.10464 <.0001	0.11373 <.0001
FL23 FLAVONO-Myricetin	0.25781 <.0001	0.52070 <.0001	0.26414 <.0001	0.25782 <.0001	0.25782 <.0001	0.25782 <.0001	0.07607 <.0001	0.03259 0.0109	0.03629 0.0046
FL24 FLAVONO-Quercetin	0.22408 <.0001	0.62275 <.0001	0.25099 <.0001	0.22408 <.0001	0.22408 <.0001	0.22408 <.0001	0.13680 <.0001	0.17754 <.0001	0.17158 <.0001

FL25 FLAVAN3-Theaflavin-3,3'-digallate	1.00000 <.0001	0.22076 <.0001	0.99212 <.0001	1.00000 <.0001	1.00000 <.0001	1.00000 <.0001	0.09964 <.0001	0.01519 0.2354	0.01934 0.1307
FL26 FLAVAN3-Theaflavin-3'-gallate	1.00000 <.0001	0.22076 <.0001	0.99212 <.0001	1.00000 <.0001	1.00000 <.0001	1.00000 <.0001	0.09964 <.0001	0.01519 0.2354	0.01934 0.1307
FL27 FLAVAN3-Theaflavin-3-gallate	1.00000 <.0001	0.22076 <.0001	0.99212 <.0001	1.00000 <.0001	1.00000 <.0001	1.00000 <.0001	0.09964 <.0001	0.01519 0.2354	0.01934 0.1307
FL28 FLAVAN3-(+)-Gallocatechin	1.00000 <.0001	0.22076 <.0001	0.99212 <.0001	1.00000 <.0001	1.00000 <.0001	1.00000 <.0001	0.09964 <.0001	0.01519 0.2354	0.01934 0.1307
FL29 Proanthocyanidin-Monomers	0.20714 <.0001	0.82386 <.0001	0.21362 <.0001	0.20713 <.0001	0.20713 <.0001	0.20713 <.0001	0.02357 0.0655	0.02909 0.0230	0.00788 0.5379
FL30 Proanthocyanidin-Dimers	0.07198 <.0001	0.86773 <.0001	0.08209 <.0001	0.07197 <.0001	0.07197 <.0001	0.07197 <.0001	0.01686 0.1876	0.05317 <.0001	0.02833 0.0268
FL31 Proanthocyanidin-Trimers	0.07363 <.0001	0.75550 <.0001	0.09984 <.0001	0.07362 <.0001	0.07362 <.0001	0.07362 <.0001	0.12882 <.0001	0.21864 <.0001	0.19303 <.0001
FL32 Proanthocyanidin-4-6mers	0.04757 0.0002	0.80366 <.0001	0.07771 <.0001	0.04756 0.0002	0.04756 0.0002	0.04756 0.0002	0.12387 <.0001	0.22661 <.0001	0.19938 <.0001
FL33 Proanthocyanidin-7-10mers	0.04836 0.0002	0.81520 <.0001	0.08022 <.0001	0.04835 0.0002	0.04835 0.0002	0.04835 0.0002	0.11527 <.0001	0.21463 <.0001	0.18937 <.0001
FL34 Proanthocyanidin-Polymers	0.04109 0.0013	0.73313 <.0001	0.09713 <.0001	0.04108 0.0013	0.04108 0.0013	0.04108 0.0013	0.13077 <.0001	0.24003 <.0001	0.21176 <.0001

	FL19	**FL20**	**FL21**	**FL22**	**FL23**	**FL24**	**FL25**	**FL26**	**FL27**
FL1 ISO-DAIDZEINA	-0.00661 0.6056	0.00087 0.9459	0.00730 0.5685	0.00600 0.6394	-0.00650 0.6114	0.02933 0.0219	-0.00846 0.5084	-0.00846 0.5084	-0.00846 0.5084
FL2 ISO-GENISTEIN	-0.00194 0.8797	-0.00040 0.9749	0.00857 0.5030	0.00421 0.7425	-0.00557 0.6632	0.00847 0.5082	-0.00859 0.5023	-0.00859 0.5023	-0.00859 0.5023
FL3 ANTHO-Cyanidin	0.16096 <.0001	0.12492 <.0001	0.31243 <.0001	0.16465 <.0001	0.08360 <.0001	0.26715 <.0001	-0.00548 0.6686	-0.00548 0.6686	-0.00548 0.6686
FL4 ANTHO-Delphinidin	0.10143 <.0001	0.00059 0.9633	0.09379 <.0001	0.11305 <.0001	0.60236 <.0001	0.26833 <.0001	-0.02112 0.0988	-0.02112 0.0988	-0.02112 0.0988
FL5 ANTHO-Malvidin	0.10146 <.0001	0.00061 0.9619	0.09379 <.0001	0.11308 <.0001	0.60238 <.0001	0.26834 <.0001	-0.02110 0.0992	-0.02110 0.0992	-0.02110 0.0992
FL6 ANTHO-Pelargonidin	0.05472 <.0001	0.08293 <.0001	0.08206 <.0001	0.08925 <.0001	0.00971 0.4483	0.12738 <.0001	-0.01078 0.3996	-0.01078 0.3996	-0.01078 0.3996
FL7 ANTHO-Peonidin	0.10896 <.0001	0.01134 0.3758	0.10429 <.0001	0.12486 <.0001	0.60330 <.0001	0.28493 <.0001	-0.02227 0.0819	-0.02227 0.0819	-0.02227 0.0819
FL8 ANTHO-Petunidin	0.10144 <.0001	0.00058 0.9641	0.09377 <.0001	0.11305 <.0001	0.60236 <.0001	0.26832 <.0001	-0.02111 0.0990	-0.02111 0.0990	-0.02111 0.0990
FL9 FLAVAN3-(+)-Catechin	0.18733 <.0001	0.07687 <.0001	0.12383 <.0001	0.21774 <.0001	0.63963 <.0001	0.37464 <.0001	0.09506 <.0001	0.09506 <.0001	0.09506 <.0001

FL29 Proanthocyanidin-Monomers	0.26703 <.0001	0.14540 <.0001	0.13724 <.0001	0.29925 <.0001	0.63041 <.0001	0.43056 <.0001	0.20713 <.0001	0.20713 <.0001	0.20713 <.0001
FL30 Proanthocyanidin-Dimers	0.20539 <.0001	0.10627 <.0001	0.14489 <.0001	0.22359 <.0001	0.60279 <.0001	0.45958 <.0001	0.07197 <.0001	0.07197 <.0001	0.07197 <.0001
FL31 Proanthocyanidin-Trimers	0.25836 <.0001	0.20895 <.0001	0.17308 <.0001	0.22776 <.0001	0.24898 <.0001	0.55183 <.0001	0.07362 <.0001	0.07362 <.0001	0.07362 <.0001
FL32 Proanthocyanidin-4-6mers	0.28276 <.0001	0.21537 <.0001	0.19611 <.0001	0.23158 <.0001	0.28007 <.0001	0.58719 <.0001	0.04756 0.0002	0.04756 0.0002	0.04756 0.0002
FL33 Proanthocyanidin-7-10mers	0.29838 <.0001	0.22224 <.0001	0.20154 <.0001	0.23476 <.0001	0.26600 <.0001	0.60404 <.0001	0.04835 0.0002	0.04835 0.0002	0.04835 0.0002
FL34 Proanthocyanidin-Polymers	0.36692 <.0001	0.23924 <.0001	0.22904 <.0001	0.24036 <.0001	0.28961 <.0001	0.56550 <.0001	0.04108 0.0013	0.04108 0.0013	0.04108 0.0013

	FL28	**FL29**	**FL30**	**FL31**	**FL32**	**FL33**	**FL34**
FL1 ISO-DAIDZEINA	-0.00846 0.5084	-0.00530 0.6785	0.00155 0.9035	0.04046 0.0016	0.04170 0.0011	0.04539 0.0004	0.03587 0.0051
FL2 ISO-GENISTEIN	-0.00859 0.5023	-0.01185 0.3546	-0.01073 0.4019	0.00516 0.6871	0.00594 0.6428	0.00701 0.5838	0.01013 0.4286
FL3 ANTHO-Cyanidin	-0.00548 0.6686	0.12265 <.0001	0.13388 <.0001	0.23331 <.0001	0.26208 <.0001	0.25419 <.0001	0.33152 <.0001
FL4 ANTHO-Delphinidin	-0.02112 0.0988	0.90762 <.0001	0.90819 <.0001	0.26243 <.0001	0.31610 <.0001	0.29333 <.0001	0.32649 <.0001
FL5 ANTHO-Malvidin	-0.02110 0.0992	0.90763 <.0001	0.90819 <.0001	0.26244 <.0001	0.31612 <.0001	0.29334 <.0001	0.32651 <.0001
FL6 ANTHO-Pelargonidin	-0.01078 0.3996	0.01652 0.1968	0.02627 0.0401	0.17204 <.0001	0.18929 <.0001	0.17983 <.0001	0.24344 <.0001
FL7 ANTHO-Peonidin	-0.02227 0.0819	0.90909 <.0001	0.91092 <.0001	0.28491 <.0001	0.34091 <.0001	0.31698 <.0001	0.35854 <.0001
FL8 ANTHO-Petunidin	-0.02111 0.0990	0.90762 <.0001	0.90818 <.0001	0.26241 <.0001	0.31609 <.0001	0.29332 <.0001	0.32648 <.0001
FL9 FLAVAN3-(+)-Catechin	0.09506 <.0001	0.94713 <.0001	0.94521 <.0001	0.37760 <.0001	0.43394 <.0001	0.41364 <.0001	0.45822 <.0001
FL10 FLAVAN3-(-)-Epigallocatechin	1.00000 <.0001	0.20714 <.0001	0.07198 <.0001	0.07363 <.0001	0.04757 0.0002	0.04836 0.0002	0.04109 0.0013
FL11 FLAVAN3-(-)-Epicatechin	0.22076 <.0001	0.82386 <.0001	0.86773 <.0001	0.75550 <.0001	0.80366 <.0001	0.81520 <.0001	0.73313 <.0001
FL12 FLAVAN3-(-)-Epicatechin 3-gallate	0.99212 <.0001	0.21362 <.0001	0.08209 <.0001	0.09984 <.0001	0.07771 <.0001	0.08022 <.0001	0.09713 <.0001

FL13 FLAVAN3-(-)-Epigallocatechin 3-gallate	1.00000 <.0001	0.20713 <.0001	0.07197 <.0001	0.07362 <.0001	0.04756 0.0002	0.04835 0.0002	0.04108 0.0013
FL14 FLAVAN3-Theaflavin	1.00000 <.0001	0.20713 <.0001	0.07197 <.0001	0.07362 <.0001	0.04756 0.0002	0.04835 0.0002	0.04108 0.0013
FL15 FLAVAN3-Thearubigins	1.00000 <.0001	0.20713 <.0001	0.07197 <.0001	0.07362 <.0001	0.04756 0.0002	0.04835 0.0002	0.04108 0.0013
FL16 FLAVA-Eriodictyol	0.09964 <.0001	0.02357 0.0655	0.01686 0.1876	0.12882 <.0001	0.12387 <.0001	0.11527 <.0001	0.13077 <.0001
FL17 FLAVA-Hesperetin	0.01519 0.2354	0.02909 0.0230	0.05317 <.0001	0.21864 <.0001	0.22661 <.0001	0.21463 <.0001	0.24003 <.0001
FL18 FLAVA-naringenin	0.01934 0.1307	0.00788 0.5379	0.02833 0.0268	0.19303 <.0001	0.19938 <.0001	0.18937 <.0001	0.21176 <.0001
FL19 FLAVONE-Apigenin	0.52753 <.0001	0.26703 <.0001	0.20539 <.0001	0.25836 <.0001	0.28276 <.0001	0.29838 <.0001	0.36692 <.0001
FL20 FLAVONE-Luteolin	0.39969 <.0001	0.14540 <.0001	0.10627 <.0001	0.20895 <.0001	0.21537 <.0001	0.22224 <.0001	0.23924 <.0001
FL21 FLAVONO-Isorhamnetin	0.03827 0.0028	0.13724 <.0001	0.14489 <.0001	0.17308 <.0001	0.19611 <.0001	0.20154 <.0001	0.22904 <.0001
FL22 FLAVONO-Kaempferol	0.67439 <.0001	0.29925 <.0001	0.22359 <.0001	0.22776 <.0001	0.23158 <.0001	0.23476 <.0001	0.24036 <.0001
FL23 FLAVONO-Myricetin	0.25782 <.0001	0.63041 <.0001	0.60279 <.0001	0.24898 <.0001	0.28007 <.0001	0.26600 <.0001	0.28961 <.0001
FL24 FLAVONO-Quercetin	0.22408 <.0001	0.43056 <.0001	0.45958 <.0001	0.55183 <.0001	0.58719 <.0001	0.60404 <.0001	0.56550 <.0001
FL25 FLAVAN3-Theaflavin-3,3'-digallate	1.00000 <.0001	0.20713 <.0001	0.07197 <.0001	0.07362 <.0001	0.04756 0.0002	0.04835 0.0002	0.04108 0.0013
FL26 FLAVAN3-Theaflavin-3'-gallate	1.00000 <.0001	0.20713 <.0001	0.07197 <.0001	0.07362 <.0001	0.04756 0.0002	0.04835 0.0002	0.04108 0.0013
FL27 FLAVAN3-Theaflavin-3-gallate	1.00000 <.0001	0.20713 <.0001	0.07197 <.0001	0.07362 <.0001	0.04756 0.0002	0.04835 0.0002	0.04108 0.0013
FL28 FLAVAN3-(+)-Gallocatechin	1.00000	0.20713 <.0001	0.07197 <.0001	0.07362 <.0001	0.04756 0.0002	0.04835 0.0002	0.04108 0.0013
FL29 Proanthocyanidin-Monomers	0.20713 <.0001	1.00000	0.98396 <.0001	0.58204 <.0001	0.60685 <.0001	0.57199 <.0001	0.59854 <.0001
FL30 Proanthocyanidin-Dimers	0.07197 <.0001	0.98396 <.0001	1.00000	0.63606 <.0001	0.67168 <.0001	0.64423 <.0001	0.64972 <.0001
FL31 Proanthocyanidin-Trimers	0.07362 <.0001	0.58204 <.0001	0.63606 <.0001	1.00000	0.99023 <.0001	0.97349 <.0001	0.93216 <.0001
FL32 Proanthocyanidin-4-6mers	0.04756 0.0002	0.60685 <.0001	0.67168 <.0001	0.99023 <.0001	1.00000	0.99329 <.0001	0.95214 <.0001

FL33 Proanthocyanidin-7-10mers	0.04835 0.0002	0.57199 <.0001	0.64423 <.0001	0.97349 <.0001	0.99329 <.0001	1.00000	0.94771 <.0001
FL34 Proanthocyanidin-Polymers	0.04108 0.0013	0.59854 <.0001	0.64972 <.0001	0.93216 <.0001	0.95214 <.0001	0.94771 <.0001	1.00000

Figure 1 (Appendix 3) - Multiple logistic regression-derived odds ratios (ORs) and corresponding 95% confidence intervals (CI) for in 1953 cases of colorectal cancer and 4154 controls, according to the highest versus the lowest quintiles of selected classes of flavonoids (with significant inverse association). Italy, 1992-1996.

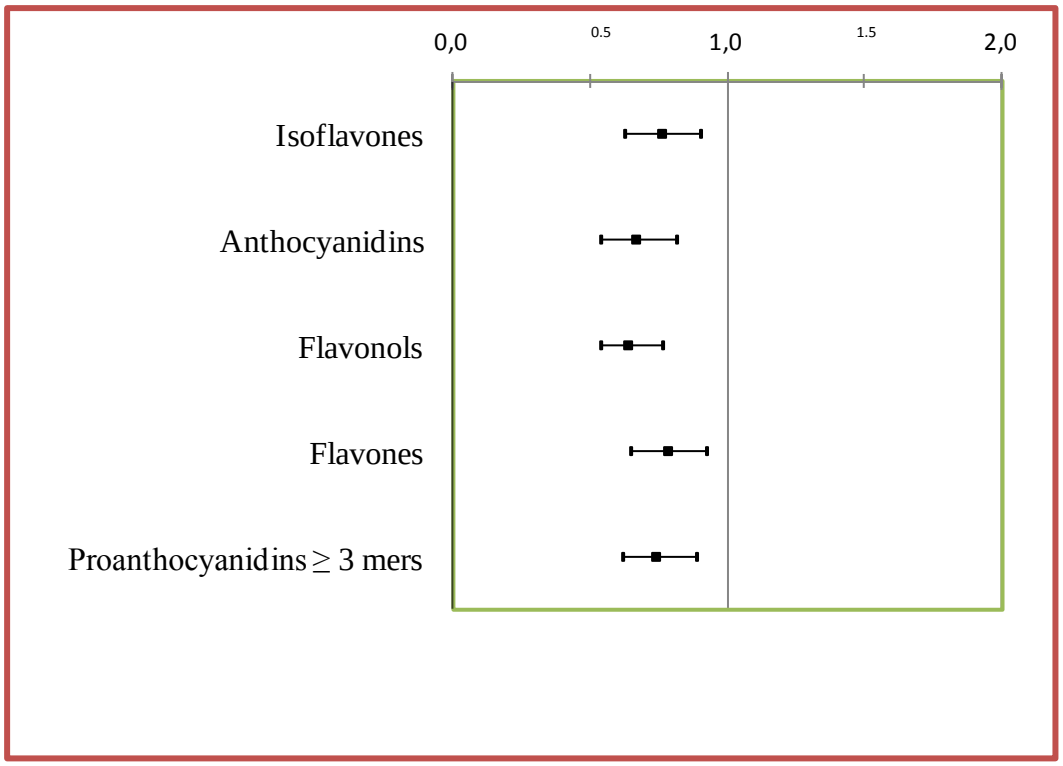


Figure 2 (Appendix 3) - Percentages of intake deriving from different sources for each flavonoids among all subjects of colorectal cancer studies. Italy, 1992-1996

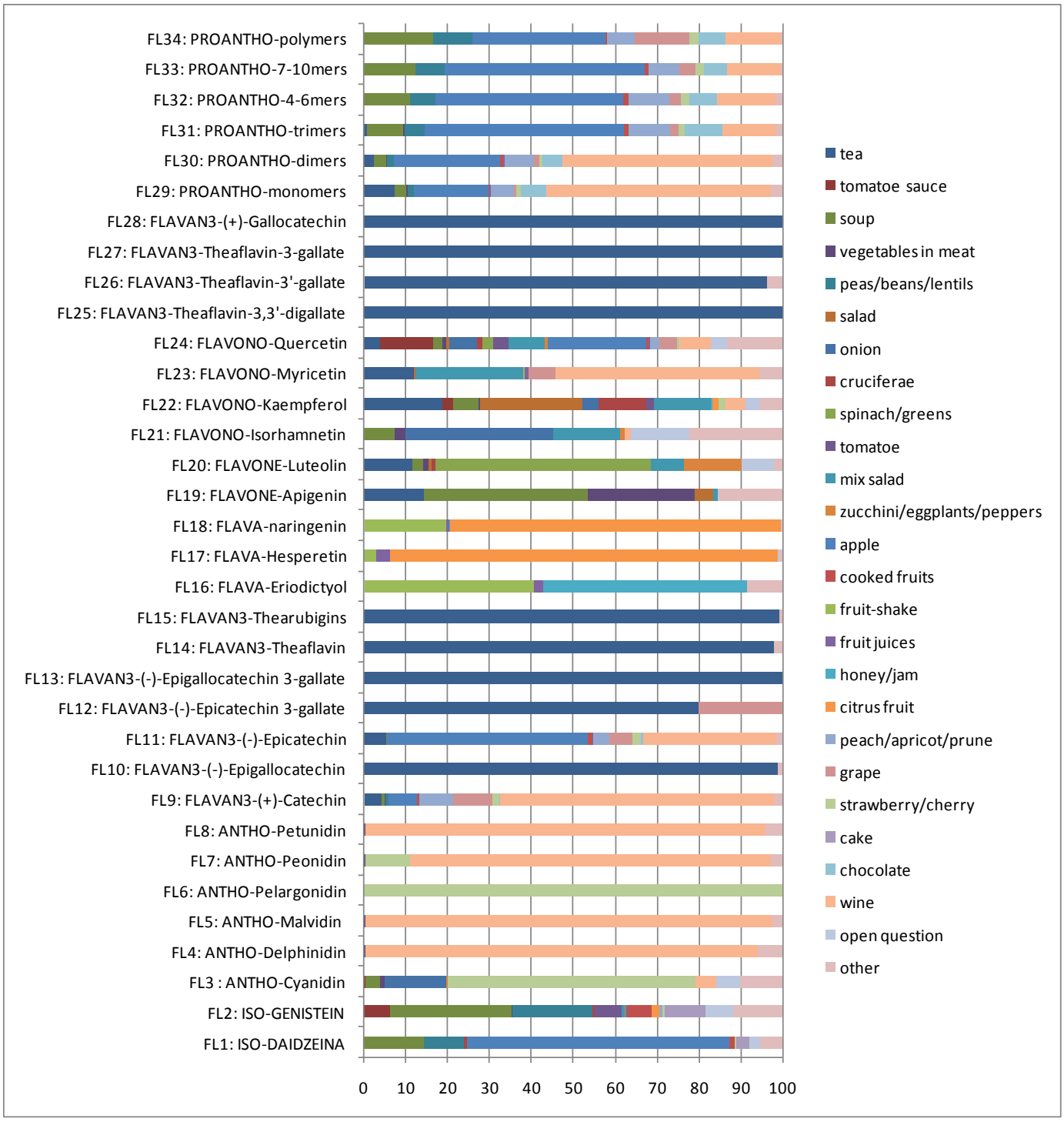


Figure 3 (Appendix 3) - Percentages of intake deriving from different sources for each flavonoids by group inserted in FA among all subjects of colorectal cancer studies. Italy, 1992-1996 .

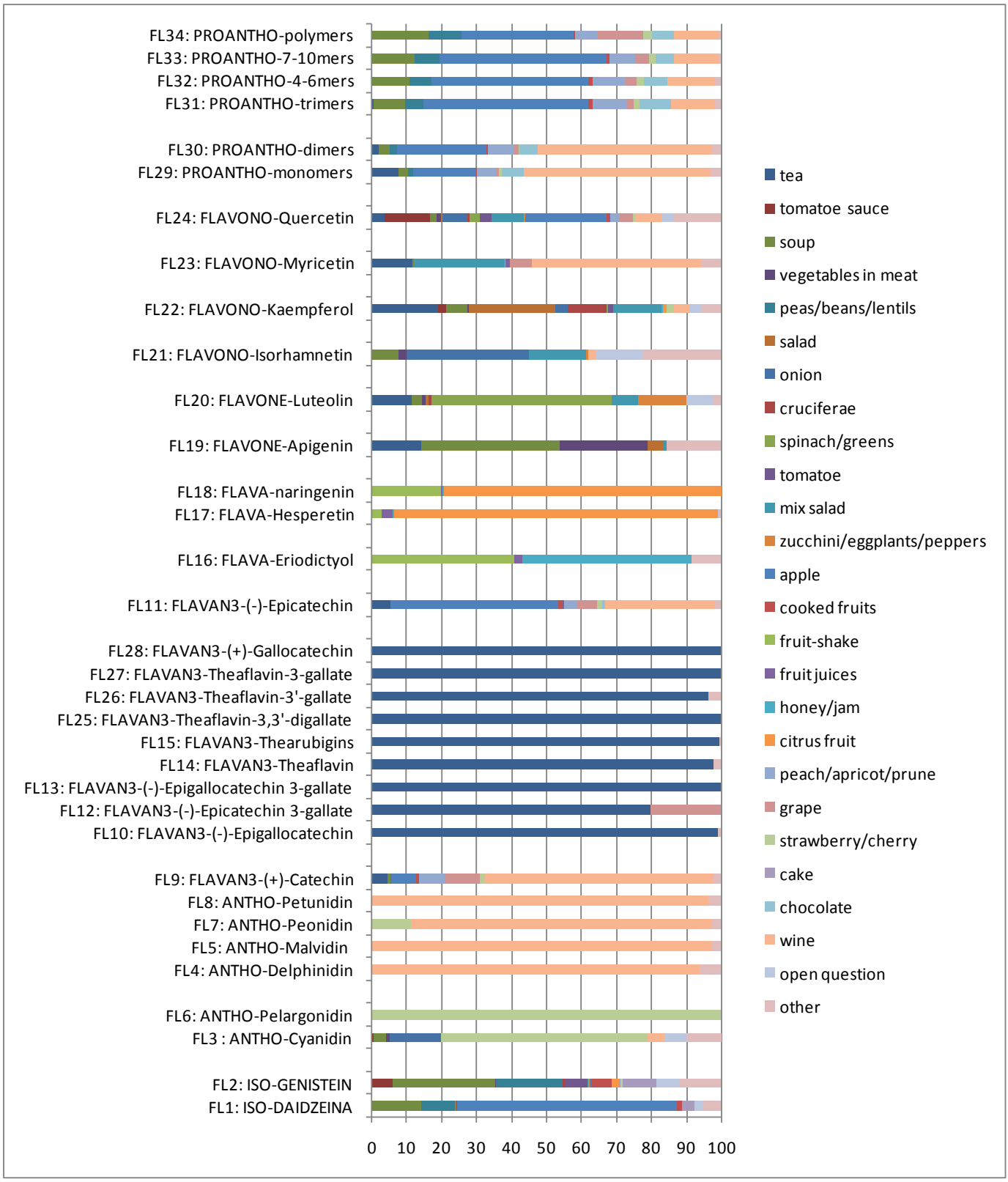


Figure 4 (Appendix 3) - Percentages of intake deriving from different sources for each flavonoids by group inserted in FA and corresponding factor (from FA with 4 factors) among all subjects of colorectal cancer studies. Italy, 1992-1996

Figure 4 – a) factor 1

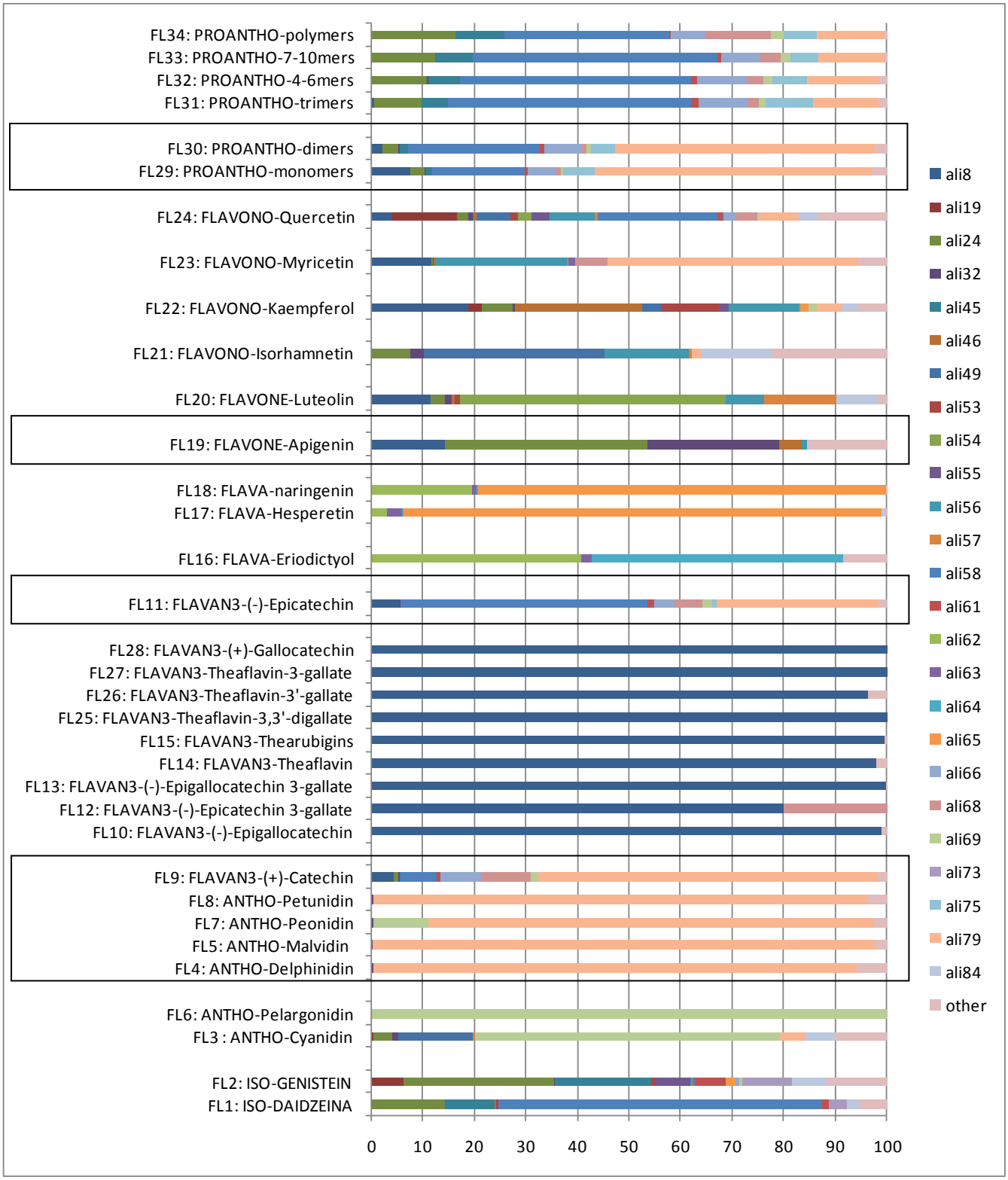


Figure 4 – b) factor 2

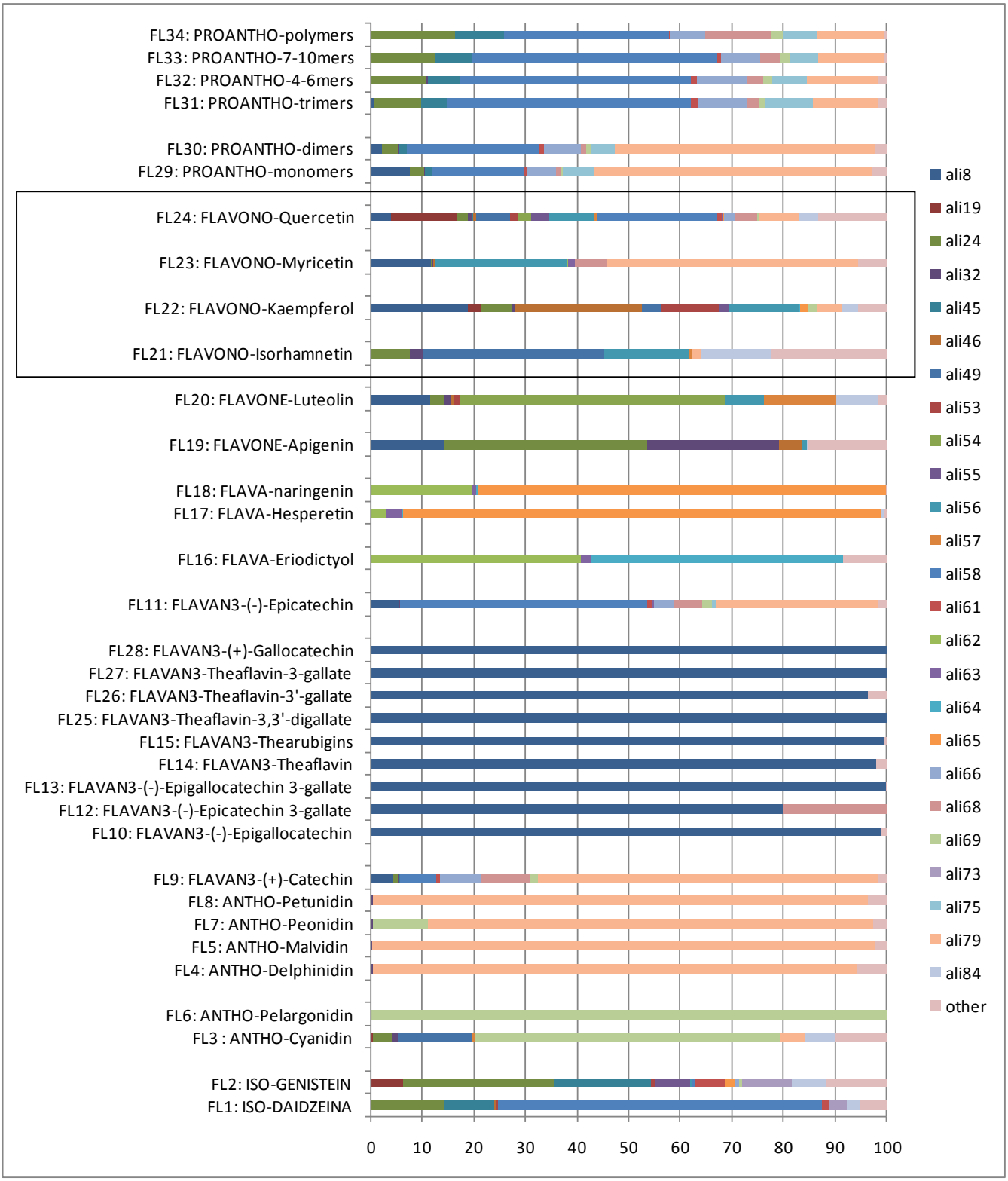


Figure 4 – c) factor 3

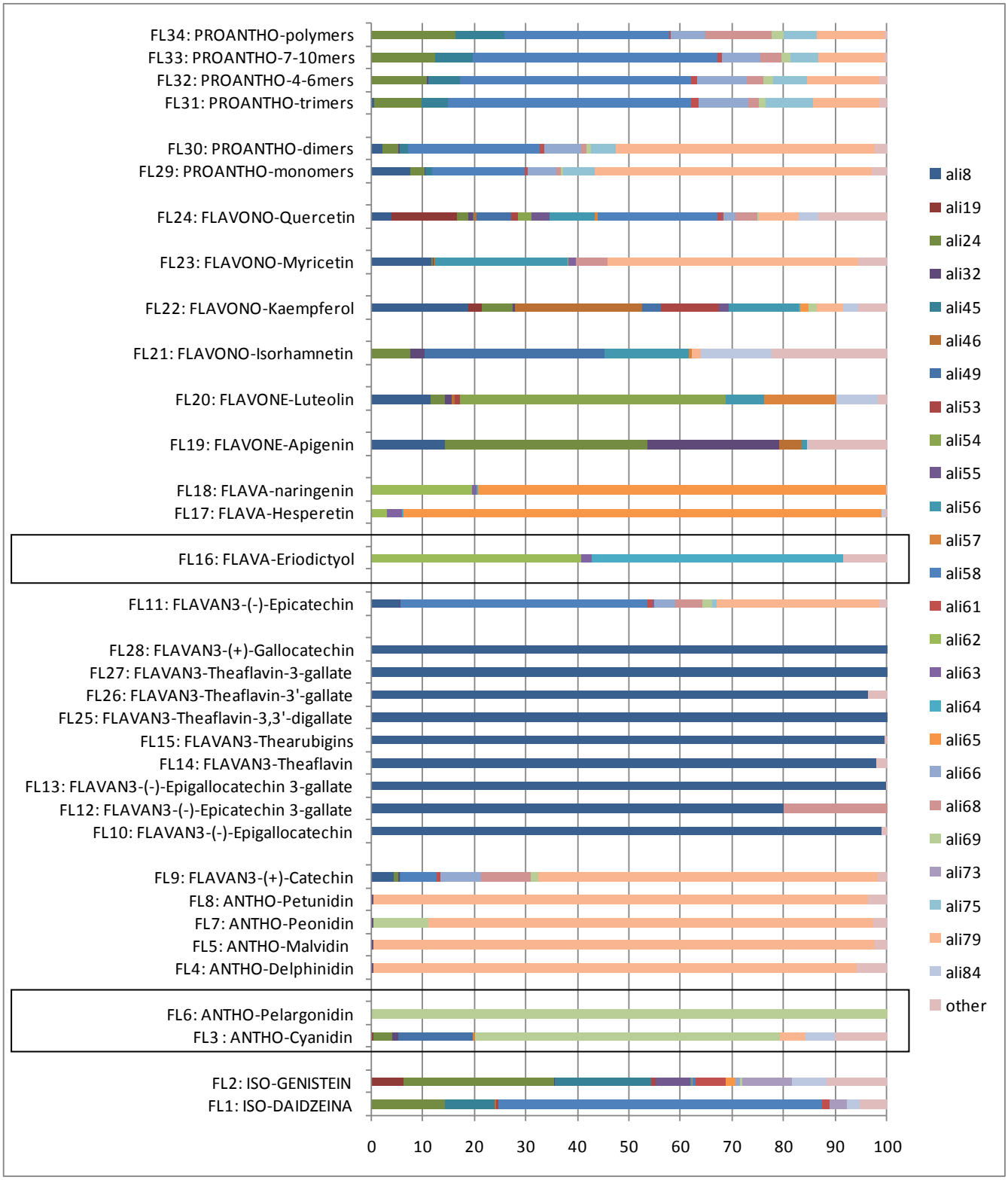


Figure 4 – d) factor 4

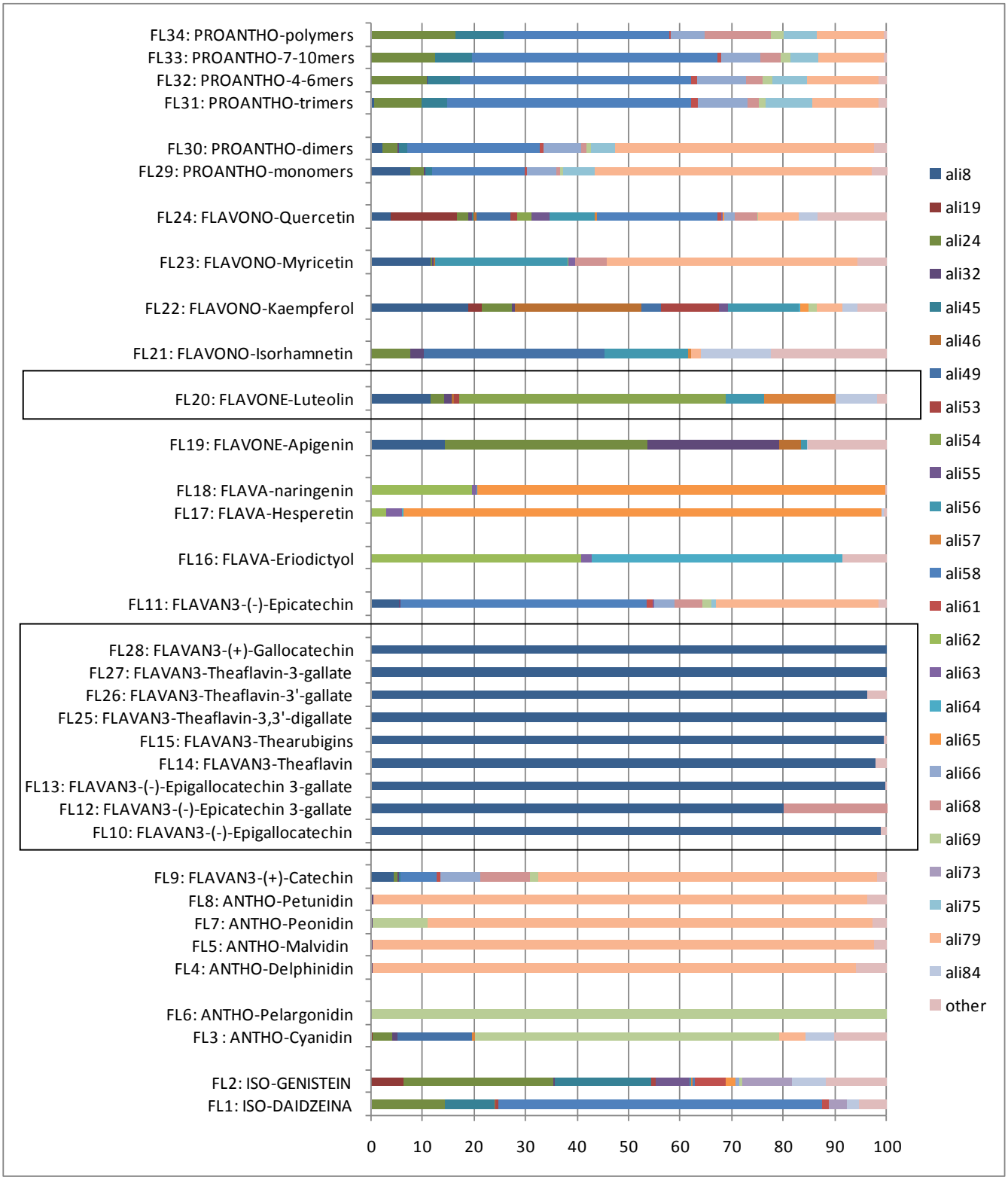


Figure 5 (Appendix 3) - Multiple logistic regression-derived odds ratios (ORs)[a] and corresponding 95% confidence intervals (CI) for colorectal cancer, according to the second and third tertiles (factor*i*t2 and factor*i*t3 for *i*=1,…,5) versus the lowest tertile of factors. Italy, 1992-1996.

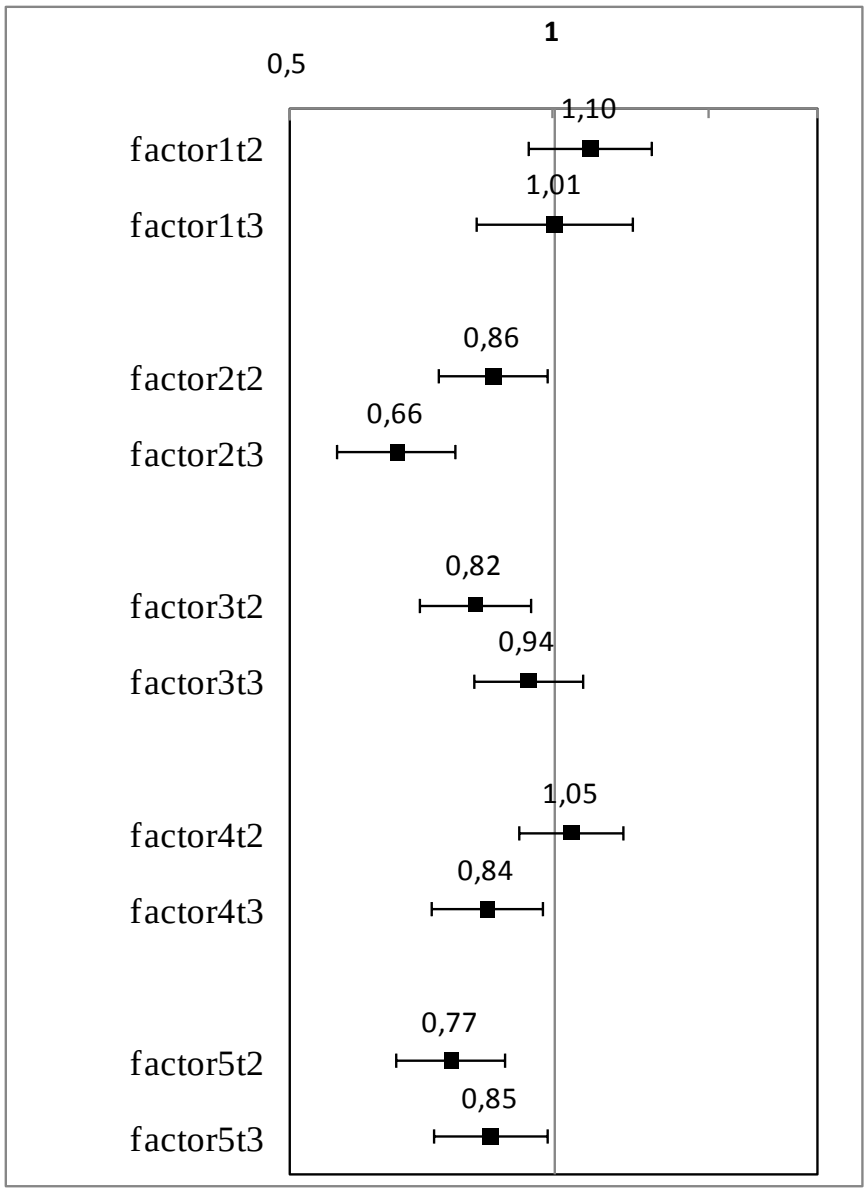


Figure 6 (Appendix 3) - Mean Intake of proanthocyanidins, monomers and dimers combined and polymers with three or more mers, deriving by wine and by other sources among of 805 oral cavity and pharyngeal cancer cases and 2,081. Italy, 1992-2005

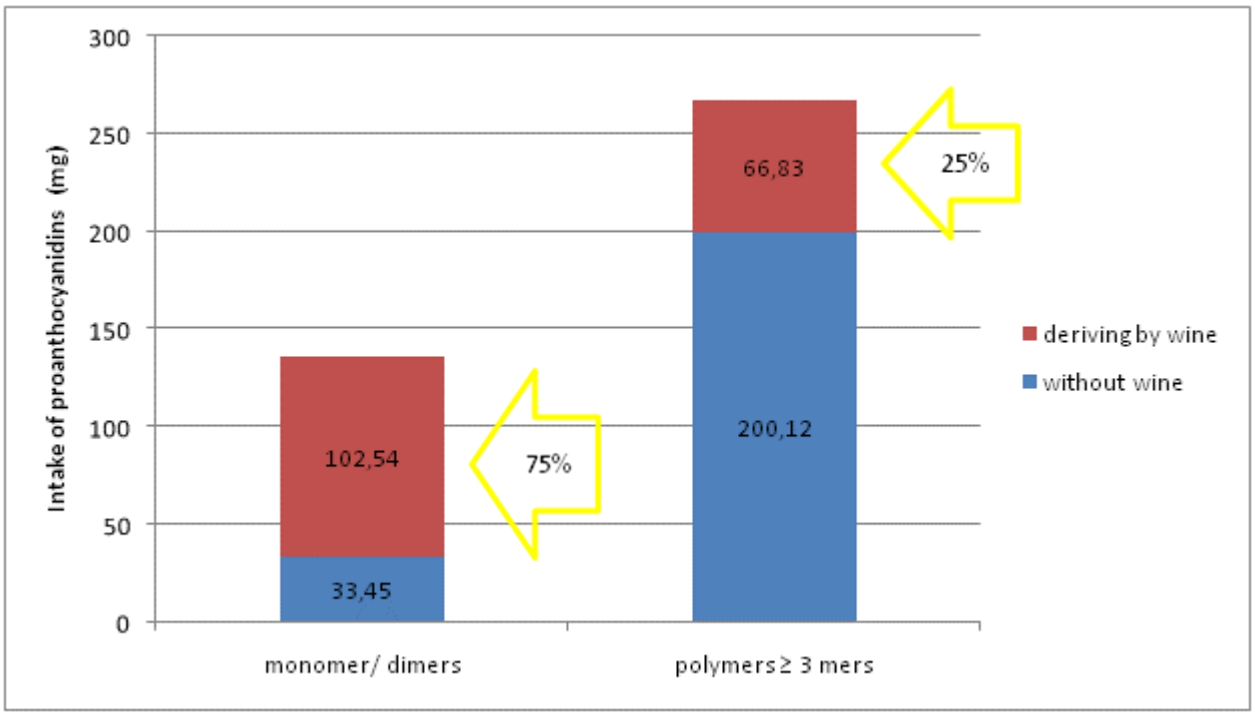


Figure 7 (Appendix 3) - Multiple logistic regression-derived odds ratios (ORs) and corresponding 95% confidence intervals (CI) for oral cavity and pharyngeal cancer, according to the highest versus the lowest tertile of monomer and dimer proanthocyanidins combined and proanthocyanidins with three or more mers, deriving by wine and without intake from wine. Italy, 1992-2005.

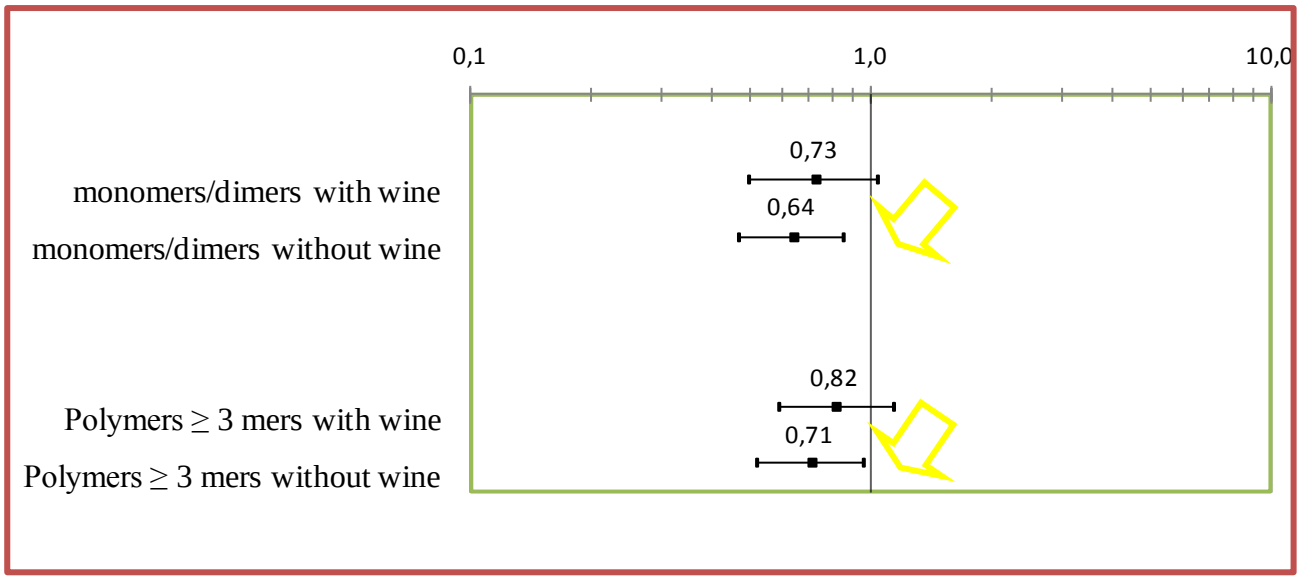


Figure 8 (Appendix 3) - Multiple logistic regression-derived odds ratios (ORs) and corresponding 95% confidence intervals (CI) for oral cavity and pharyngeal cancer, according to the second and third tertiles (factor*i*t2 and factor*i*t3 for *i*=1,2,3) versus the lowest tertile of factors. Italy, 1992-2005.

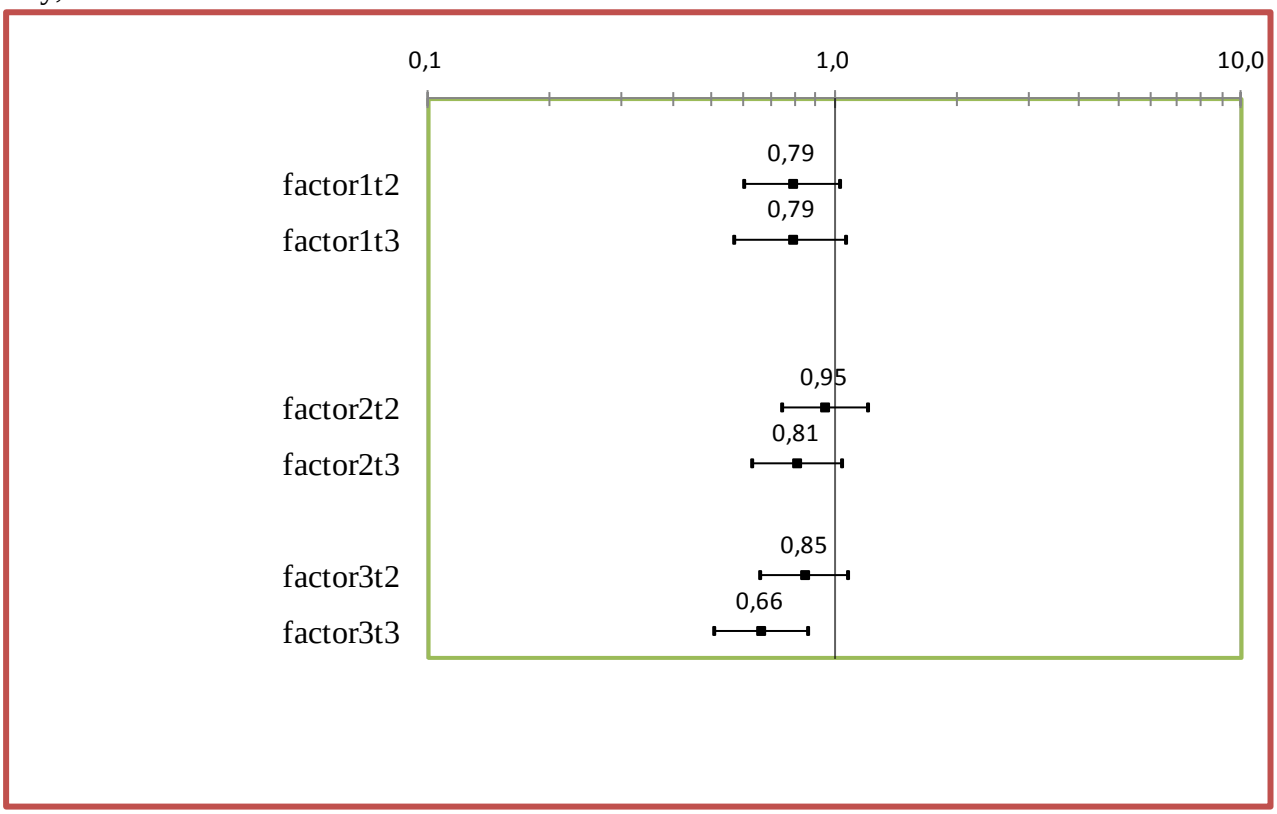
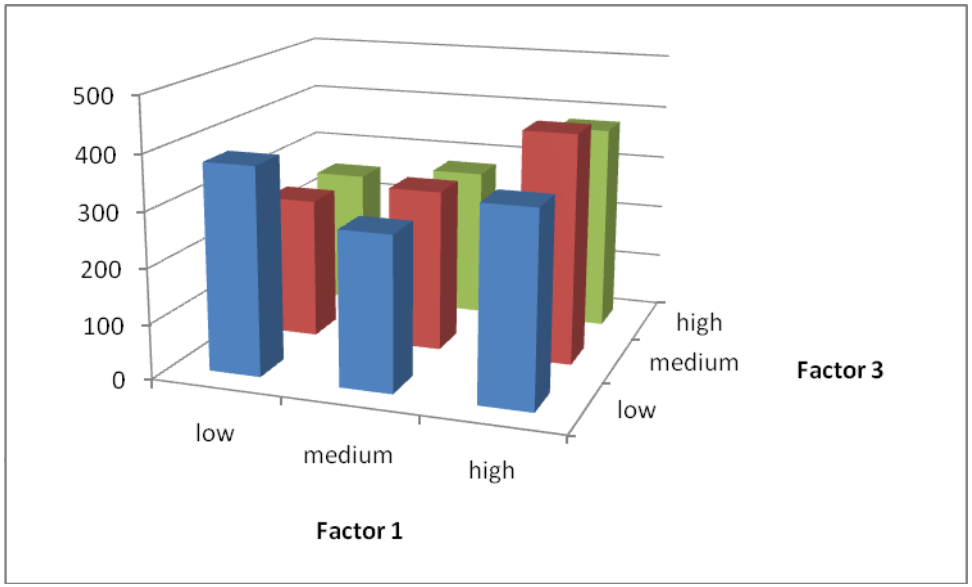


Figure 9 (Appendix 3) a), b) - Distribution of 2886 subjects from oral and pharyngeal cancer study according to levels of factor 1 and factor 3 and levels of alcohol habits and factor 3. Italy, 1992-2005.

a)



b)

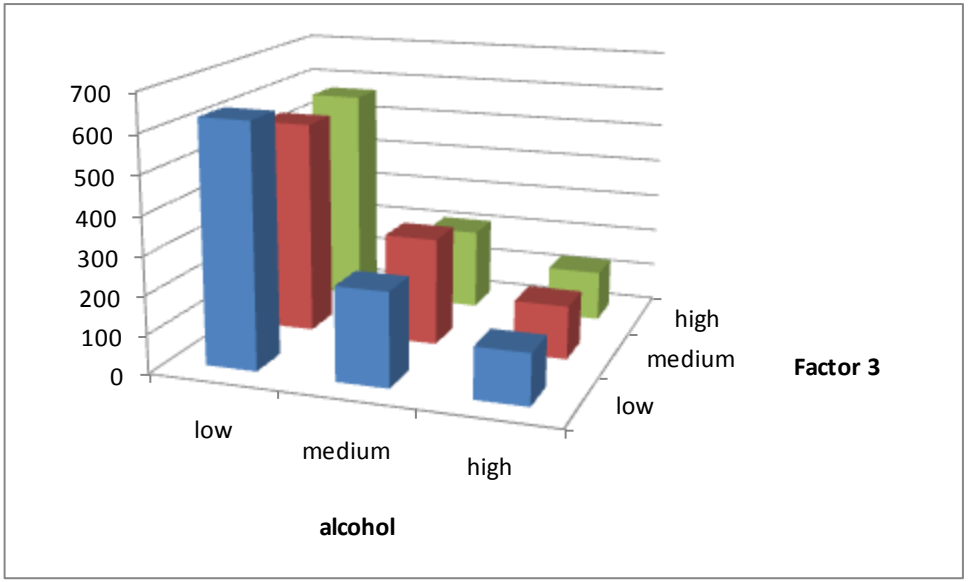
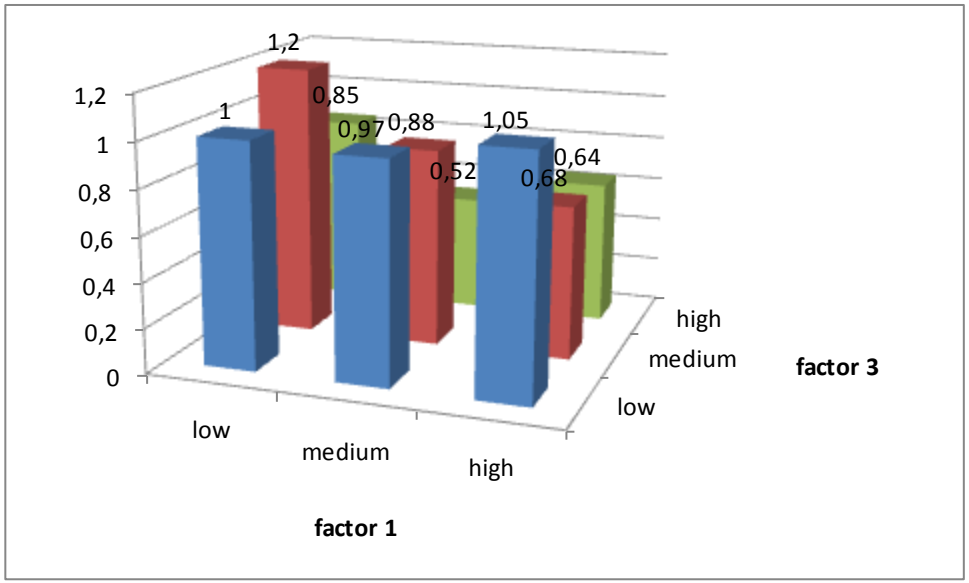


Figure 10 (Appendix 3) a), b) - Multiple logistic regression-derived odds ratios (ORs) for oral cavity and pharyngeal cancer, according to highest versus the lowest level of factor 1 and factor 3, and of alcohol habits and factor 3 . Italy, 1992-2005.

a)



b)

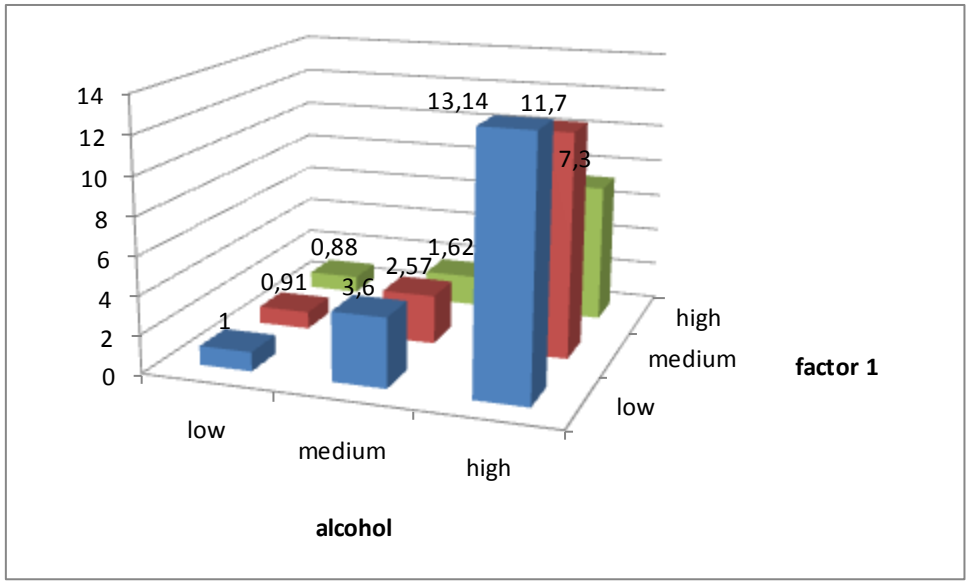


Figure 11 (Appendix 3) - Multiple logistic regression-derived odds ratios (ORs) and corresponding 95% confidence intervals (CI) for oral cavity and pharyngeal cancer, according to the highest versus the lowest quintile of flavonoid intake and residuals of flavonoids on FRAP. Italy, 1992-2005.

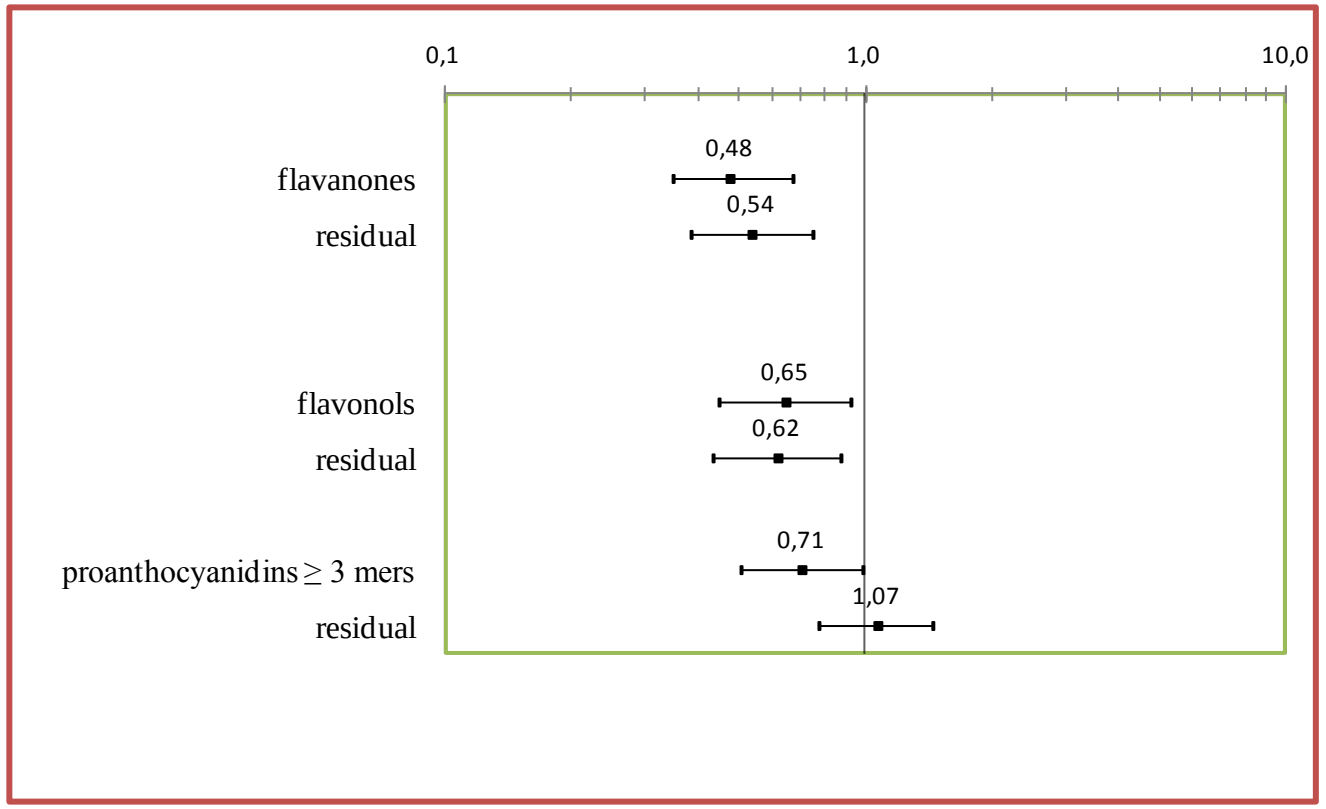
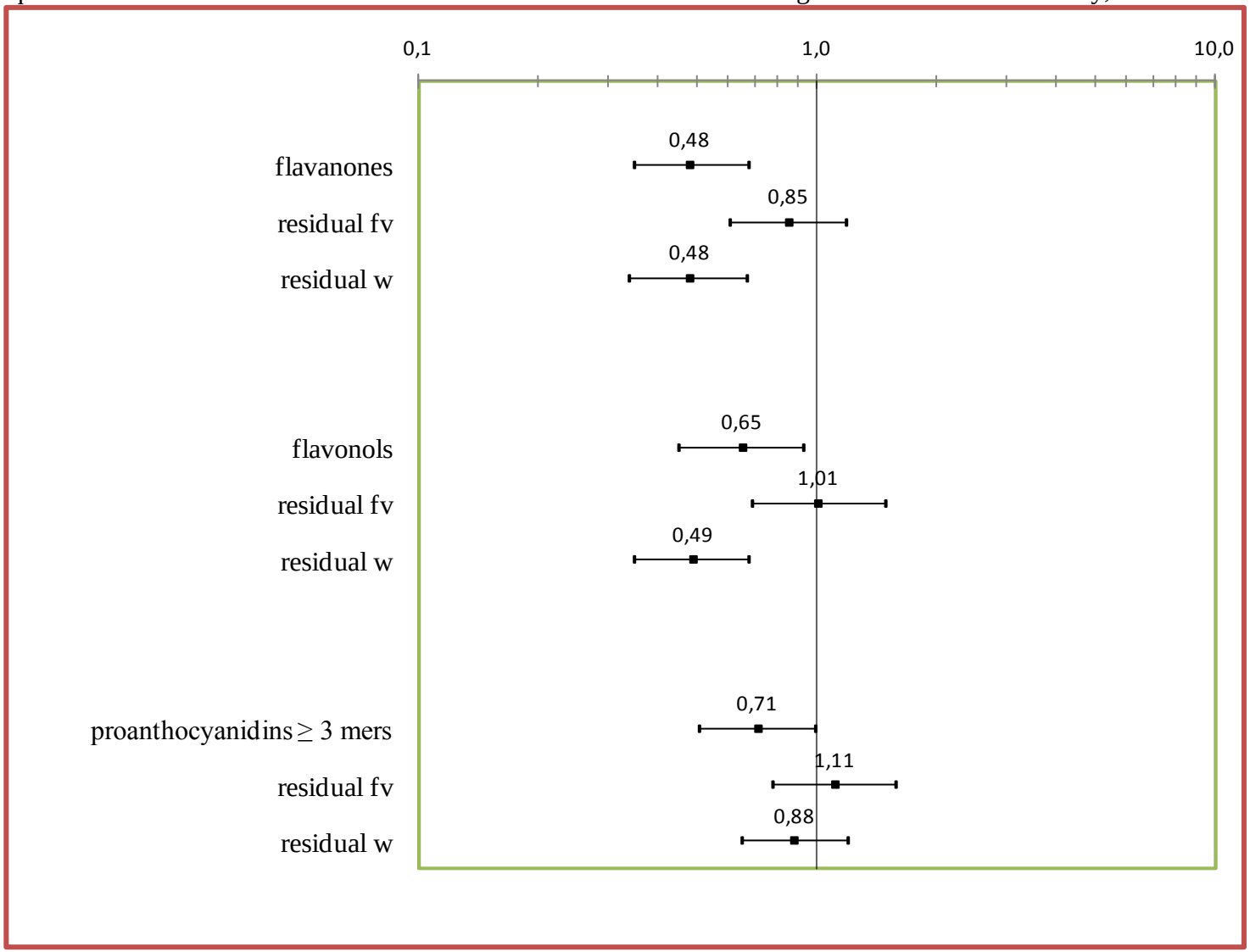


Figure 1 (Appendix 4) - Multiple logistic regression-derived odds ratios (ORs) and corresponding 95% confidence intervals (CI) for oral cavity and pharyngeal cancer, according to the highest versus the lowest quintile of flavonoid intake and residuals of flavonoids on fruit and vegetables and on water. Italy, 1992-2005.



APPENDIX 4 – SAS programs

```
data a;
set die.colon;
run;
proc sort data=a;
by v2;
run;
ods rtf file='C:\Users\mrossi\Desktop\TESI\OUTPUT\means1.doc';
proc means data=a MAXDEC=2;
var fl1-fl34;
by v2;
run;
ods rtf close;

proc univariate PLOT;
var fl1-fl34;
run;
proc univariate normaltest;;
var fl1-fl34;
where v2=2;
run;
proc univariate normaltest; PROBPLOT fl1/normal (mu=est sigma=est) ;
var fl1-fl34;
run;
proc univariate data=a normaltest; QQPLOT fl1/normal (mu=est sigma=est) ;
var fl1-fl34;histogram fl1-fl34;
run;

ods rtf file='C:\Users\mrossi\Desktop\TESI\OUTPUT\corr1.doc';
proc corr data=a ;
var fl1-fl34;
run;
ods rtf close;


proc standard data=a mean=0 std=1 out=stand;
var fl1-fl34 ali1-ali78 pro prop prott;
run;

data b (keep=v1 v2 v4 fl1-fl34 ali1-ali78 pro prop prott);
set stand;
run;


ods rtf file='C:\Users\mrossi\Desktop\TESI\OUTPUT\factor1.doc';
title1 'Analisi fattoriale, rotazione=varimax, metodo=principal (...)';
proc factor data=b method=principal corr score res /*out=primaf*/ msa
mineigen=1 priors=one /*n=1*/ r=varimax scree ;
var fl1-fl34;
run;
ods rtf close;


proc factor data=b method=principal corr score res /*out=primaf*/ msa
mineigen=1   priors=one n=3 r=varimax scree;
var fl3 fl9 fl10 fl16 fl17 fl19 fl20-fl24 fl29 fl31;
run;/* MSA TOT = 0.55*/

proc factor data=b method=principal corr score res /*out=primaf*/ msa
mineigen=1   priors=one n=3 r=varimax scree;
var fl3 fl4 fl10 fl11 fl16 fl17 fl19-fl24 fl29 fl31;
run;/* MSA TOT= 0.47 */
```

```
proc factor data=b method=principal corr score res /*out=primaf*/ msa
mineigen=1   priors=one n=3 r=varimax scree;
var fl3 fl4 fl10 fl11 fl16 fl17 fl19-fl24 fl29 ;
run;/* MSA TOT= 0.68 */

proc factor data=b method=principal corr score res /*out=primaf*/ msa
mineigen=1   priors=one n=3 r=varimax scree;
var fl3 fl4 fl10 fl11 fl16 fl17 fl19-fl24 fl31 ;
run;/* MSA TOT= 0.60 */

proc factor data=b method=principal corr score res /*out=primaf*/ msa
mineigen=1   priors=one n=3 r=varimax scree;
var fl1 fl3 fl4 fl10 fl11 fl16 fl17 fl19-fl24 fl29 fl31;
run;/* MSA TOT= 0.47 */


data aa;
set a;
s3=fl3+fl6;
s4=fl4+fl5+fl7+fl8+fl9;
s10=fl10+fl12+fl13+fl14+fl15+fl25+fl26+fl27+fl28;
s11=fl11;
s16=fl6;
s17=fl7+fl18;
s19=fl9;
s20=fl20;
s21=fl21;
s22=fl22;
s23=fl23;
s24=fl24;
s29=fl29+fl30;
s31=fl31+fl32+fl33+fl34;
run;

proc standard data=aa mean=0 std=1 out=stand;
var fl1-fl34 ali1-ali78 pro prop prott s3 s4 s10 s11 s16 s17 s19-s24 s29
s31;
run;

data b (keep=v1 v2 v4 fl1-fl34 ali1-ali78 pro prop prott s3 s4 s10 s11 s16
s17 s19-s24 s29 s31);
set stand;
run;

ods rtf file='C:\Users\mrossi\Desktop\TESI\OUTPUT\factor1.doc';
proc factor data=b method=principal corr score res out=primaf msa
/*mineigen=1*/ /*heywood*/   priors=one n=5 r=varimax scree;
var s3 s4 s10 s11 s16 s17 s19-s24 s29 s31;
run;

proc princomp data=b score n=5 ;
var s3 s4 s10 s11 s16 s17 s19-s24 s29 s31;
run;
ods rtf close;/* MSA TOT= 0.61 */


* 4 factors ;
proc sort data=aa;
by  v1 v2 v4;
run;
proc sort data=primaf;
by  v1 v2 v4;
run;
data merge4;
merge aa primaf;
```

```
by v1 v2 v4;
run;

data aaa;
set merge4;
where v2=2;
run;

*****TERTILES***************;
%macro terz;
%let c1=factor1;
%let c2=factor2;
%let c3=factor3;
%let c4=factor4;
%do i=1 %to 4;
proc univariate data=aaa noprint;
var &&c&i;
output out=pp pctlpts=33.3 66.6
       pctlpre=pr&&c&i ;
run ;

proc print data=pp;
run;

%macro perc(nom) ;
    data _null_;
         set pp ;
         %global &nom.33_3 &nom.66_6 ;
         call symput("&nom.33_3",&nom.33_3);
         call symput("&nom.66_6",&nom.66_6);
    run;
   data _null_ ;
    %put &&&nom.33_3 ;
    %put &&&nom.66_6 ;

%mend ;


%perc(pr&&c&i) ;

%let a=pr&&c&i..33_3;
%let b=pr&&c&i..66_6;

data merge4;
   set merge4;
%dumm(&&c&i,&&c&i..t,2,&&&a,&&&b) ;
&&c&i..tfr=&&c&i..t1+2*&&c&i..t2+3*&&c&i..t3;
run;
%end;
%mend;
%terz;

proc logistic data=merge4;
model v2=factor1t2-factor1t3 factor2t2-factor2t3 factor3t2-factor3t3
factor4t2-factor4t3 etaq1-etaq3 etaq5-etaq7 sex2 cen2-cen6    edu2 edu3
nnfis2 nnfis3
famint bmi2-bmi5 alcol2-alcol4 ennoal2-ennoal5;
run;

* 5 factors;
proc factor data=b method=principal corr score res out=primaf msa
/*mineigen=1*/ /*heywood*/   priors=one n=5 r=varimax scree;
var s3 s4 s10 s11 s16 s17 s19-s24 s29 s31;
run;
```

```
proc sort data=aa;
by  v1 v2 v4;
run;
proc sort data=primaf;
by  v1 v2 v4;
run;
data merge5;
merge aa primaf;
by v1 v2 v4;
run;
data aaa;
set merge5;
where v2=2;
run;
%macro terz;
%let c1=factor1;
%let c2=factor2;
%let c3=factor3;
%let c4=factor4;
%let c5=factor5;

%do i=1 %to 5;
proc univariate data=aaa noprint;
var &&c&i;
output out=pp pctlpts=33.3 66.6
       pctlpre=pr&&c&i ;
run ;

proc print data=pp;
run;

%macro perc(nom) ;
    data _null_;
         set pp ;
         %global &nom.33_3 &nom.66_6 ;
         call symput("&nom.33_3",&nom.33_3);
         call symput("&nom.66_6",&nom.66_6);
    run;
   data _null_ ;
    %put &&&nom.33_3 ;
    %put &&&nom.66_6 ;

%mend ;


%perc(pr&&c&i) ;

%let a=pr&&c&i..33_3;
%let b=pr&&c&i..66_6;

data merge5;
   set merge5;
%dumm(&&c&i,&&c&i..t,2,&&&a,&&&b) ;
&&c&i..tfr=&&c&i..t1+2*&&c&i..t2+3*&&c&i..t3;
run;
%end;
%mend;
%terz;

proc logistic data=merge5;
model v2=factor1t2-factor1t3 factor2t2-factor2t3 factor3t2-factor3t3
factor4t2-factor4t3 factor5t2-factor5t3  etaq1-etaq3 etaq5-etaq7 sex2 cen2-
cen6    edu2 edu3 nnfis2 nnfis3
famint bmi2-bmi5 alcol2-alcol4 ennoal2-ennoal5;
ods output oddsratios = ff;
```

```
run;
proc logistic data=merge5;
model v2=factor1tfr factor2tfr factor3tfr factor4tfr factor5tfr  etaq1-
etaq3 etaq5-etaq7 sex2 cen2-cen6    edu2 edu3 nnfis2 nnfis3
famint bmi2-bmi5 alcol2-alcol4 ennoal2-ennoal5;
ods output oddsratios = ff;
run;
proc univariate data=merge5 normaltest; PROBPLOT factor1/normal (mu=est
sigma=est) ;
var factor1;histogram factor1;
run;
proc univariate data=merge5 normaltest; PROBPLOT factor2/normal (mu=est
sigma=est) ;
var factor2;histogram factor2;
run;
proc univariate data=merge5 normaltest; PROBPLOT factor3/normal (mu=est
sigma=est) ;
var factor3;histogram factor3;
run;
proc univariate data=merge5 normaltest; PROBPLOT factor4/normal (mu=est
sigma=est) ;
var factor4;histogram factor4;
run;
proc univariate data=merge5 normaltest; PROBPLOT factor5/normal (mu=est
sigma=est) ;
var factor5;histogram factor5;
run;


*****QUINTILES***************;
%macro q1;

%let c1=factor1;
%let c2=factor2;
%let c3=factor3;
%let c4=factor4;
%let c5=factor5;

%do i=1 %to 5;
proc univariate data=aaa NOPRINT ;
     var &&c&i ;
    output out=pp pctlpts=20 40 60 80
       pctlpre=pr&&c&i ;
run ;
proc print data=pp;
run;
%macro perc(nom) ;
    data _null_;
         set pp ;
         %global &nom.20 &nom.40 &nom.60 &nom.80 ;
         call symput("&nom.20",&nom.20);
         call symput("&nom.40",&nom.40);
         call symput("&nom.60",&nom.60);
          call symput("&nom.80",&nom.80);
    run;
   data _null_ ;
    %put &&&nom.20 ;
    %put &&&nom.40 ;
    %put &&&nom.60 ;
    %put &&&nom.80 ;
%mend ;

%perc(pr&&c&i) ;

%let a=pr&&c&i..20;
```

```
%let b=pr&&c&i..40;
%let c=pr&&c&i..60;
%let d=pr&&c&i..80;


data merge5;
   set merge5;
%dumm(&&c&i,&&c&i,4,&&&a,&&&b, &&&c,&&&d) ;
&&c&i..fr=&&c&i..1+2*&&c&i..2+3*&&c&i..3+
                4*&&c&i..4+5*&&c&i..5 ;
&&c&i..cq=&&c&i/(&&&d-&&&a);

run;
%end;
%mend;
%q1;

proc logistic data=merge5;
model v2=factor12-factor15 factor22-factor25 factor32-factor35 factor42-
factor45 factor52-factor55  etaq1-etaq3 etaq5-etaq7 sex2 cen2-cen6    edu2
edu3 nnfis2 nnfis3
famint bmi2-bmi5 alcol2-alcol4 ennoal2-ennoal5;
run;
```

Acknowledgment

Ringrazio i miei mentori prof. Carlo La Vecchia e prof. Dimitrios Trichopoulos.

Ringrazio il Dipartimento di Medicina del Lavoro "Clinica del Lavoro L. Devoto" Sezione di Statistica Medica e Biometria "G.A. Maccacaro".Università degli Studi di Milano, Facoltà di Medicina e Chirurgia.

Ringrazio il Dipartimento di Epidemiologia dell'Istituto di Ricerche Farmacologiche Mario Negri, un grande gruppo di lavoro che fa della collaborazione la sua forza, in particolare la dott.ssa Paola Bertuccio.

Ringrazio il prof. Adriano Decarli per l'aiuto nell'interpretazione dei risultati e la prof. Pagona Lagiou per il continuo sostegno al mio lavoro di ricerca.

Ringrazio infine la dott.ssa Annalisa Pascarella che ha dedicato a questa ricerca il suo primo giorno di lavoro.